Enhancing School Success Prediction with FRC and Merged GNN

Melike Yildiz Aktas[1 [0000-0001-9138-3630]], Aadyant Khatri[1 [0009-0000-7574-7043]], Mariam Almutairi[1 [0009-0009-3468-6091]], Lulwah Alkulaib[2 [0000-0001-9827-0882]], and Chang-Tien Lu[1 [0000-0003-3675-0199]]

[1] Department of Computer Science, Virginia Tech, Falls Church, VA 22043, USA
[2] Department of Computer Science, Kuwait University, Kuwait
{melike, aadyant, malmutairi, lalkulaib, clu}@vt.edu

Abstract. This study presents an innovative framework for predicting school success, leveraging Large Language Models (LLMs) to define ground truth labels based on comprehensive school information with Factor-Reasoning-Classification (FRC) prompting. In this article, we conceptualize school data as a complex social network and create two different graphs where schools are defined as nodes. Similarity Graph captures school similarities, integrating factors such as graduation rates, ACT scores, socioeconomic conditions, crime rates, and community resources. Geographic Proximity Graph models spatial relationships among schools using geographical coordinates. We define Merged GNN that enhances prediction accuracy by incorporating both similarity-based and spatial proximity-based information. Our approach leverages Graph Neural Networks (GNN) to predict the most probable labels the LLM model identifies. Experimental results on the school success dataset not only demonstrate the superior predictive performance of our methodology over baseline models but also highlight the importance of integrating diverse sources of information for accurate prediction and analysis.

Keywords: Data labeling · School success prediction · Graph neural network · Classification · Spatial analysis.

1 Introduction

The choice of high school profoundly impacts students' futures including college admissions and career trajectories. Yet, public school quality is often affected by external factors like geographic location and community resources [1]. Given the significant impact of these factors on educational outcomes, understanding school performance is essential. Prioritizing school achievement prediction empowers parents, urban planners, and policymakers to make informed decisions for societal benefit. A school's success hinges on several factors, which can be grouped into four categories. First, school climate like teaching environment and safety of students are pivotal [2]. Second, crime rates can directly impact the educational environment [3]. Third, social amenities such as libraries and parks enrich the learning experience [4]. Finally, socioeconomic factors like employment rates, income distribution, and community health can influence the educational landscape. Understanding these socioeconomic dynamics is crucial for enhancing educational equity and opportunity [5].

Students' prior academic achievements, demographic background, and psychological traits are widely recognized as the primary factors influencing the prediction of academic success [6, 7]. Predicting school and student success can be determined by factors such as residential crowding [8], crime rates [9, 10], socioeconomic background [11, 12], and mental [13] and physical health [14] of the community or individuals.

In our study, the prediction of school success rather than individuals is highlighted. We present an innovative approach for predicting school success by integrating two distinct graph structures. The first graph is constructed based on the similarity of schools, utilizing various parameters such as graduation rate, educational attainment levels, socioeconomic conditions, crime rates, and the availability of community resources such as parks and libraries. This graph encapsulates the multidimensional aspects of school environments and their surrounding communities, comprehensively representing school characteristics. The second graph is built using geographical coordinates (latitude and longitude) to capture the spatial proximity between schools. By exploiting the geographic proximity of different locations, this graph allows us to model the spatial dependencies, reflecting their geographical relationships.

For ground truth labeling of school success, we present LLM-based Factor-Reasoning-Classification prompting. In our method, critical factors that might impact school success are asked, encompassing both positive and negative influences, and is provided a brief explanation of the diagnostic reasoning process, explaining the rationale behind the classification of each school's success level. Drawing upon the identified factors and diagnostic reasoning, the model assigns success classes (high, medium, or low) to each school. Merged GNN is designed to predicts the success levels that our prompt model assigns. Through this comprehensive approach, our methodology aims to provide a robust framework for assessing and understanding school performance.

Experimental evaluations conducted on real-world school success dataset demonstrates the effectiveness of our proposed approach. Comparative analyses against baseline models showcase the superior predictive performance achieved by integrating multiple graph structures. Our findings underscore the importance of incorporating both similarity-based and spatial proximity-based information for accurate school success prediction, highlighting the potential of graph-based techniques in education analytics.

We summarize our contribution as follows:

- **Data labeling prompting:** We introduce a novel data labeling prompting strategy utilizing Large Language Models (LLMs), which enables the automatic definition of school success classes based on comprehensive information that reduces the need for human labeling.

- **Predictive modeling (Merged GNN) with MergeGraph:** We advance predictive modeling techniques by employing different types of message passing of GNN, a cutting-edge approach for analyzing graph-structured data.
- **Analyzing comprehensive school success datasets:** We introduce an approach of combining heteregenous data sources into a single framework, enabling a nuanced understanding of school success rather than solely focusing on student success.
- **Conducting extensive experiment:** We investigate an experiment on a real-world Chicago school dataset to demonstrate that our model is capable of capturing the shape of success classification.

2 Related Work

In this section, three groups of related work are discussed: 1) School and student success, 2) LLM-Based data classification and labeling, and 3) Node classification with graph neural networks.

2.1 School and Student Success

The discovery of factors affecting academic success is an ongoing field of study as the definition of success evolves over time. Emotional and social competencies play a pivotal role in self-awareness, higher educational attainment, improved mental and emotional health, and a decrease in engagement with criminal activities [13]. Conversely, it's evident that reductions in crime rates and improvements in overall welfare also contribute to the pathways to success. Woehr and Newman [15] demonstrated that the presence of dogs can alleviate stress among elementary and high school students, shedding light on innovative methods to support student well-being. Bastedo et al. [16] concluded that both raw and contextualized measures of high school GPA, along with ACT scores, serve as robust predictors for underrepresented students, influencing outcomes such as freshman college GPA, retention rates, and graduation success. Additionally, providing a positive school environment has been identified as a powerful intervention to improve academic performance and promote holistic student well-being by Daily et al. [2]. Rodriguez et al. [17] delved into the interaction between socioeconomic status and academic achievement, finding a positive correlation, highlighting social and economic factors in shaping educational outcomes.

Hashim et al. [7] compared the performances of several supervised machine learning algorithms, such as Decision Tree, Naïve Bayes, Logistic Regression, Support Vector Machine, K-Nearest Neighbour, Sequential Minimal Optimisation, and Neural Network for forecasting student success in bachelor study programmes. Malik and Jothimani [18] proposed a framework to choose important features that impact student quality and reduce dropout rates. Natek and Zwilling [19] conducted a study focusing on specific students, analyzing their academic records and family backgrounds to predict performance, utilizing models such as Rep Tree, J48, and M5P. Khalaf et al. [20] forecasted exam outcomes leveraging student-filled questionnaires on social activities. In the context of predicting student dropout, Yukselturk et al. [21] explored various models, including K-Nearest Neighbour, Decision Tree, Naïve Bayes, and Neural Network.

2.2 LLM-Based Data Classification and Labeling

Recently, large language models (LLMs) have drawn significant attention for tasks like text classification, data annotation, and content creation. Trained on vast text and multi-modal data, these models excel in identifying patterns and connections, enabling them to tackle various NLP tasks with exceptional adaptability and generalization [22–24].

One of the most significant benefits of LLMs is their ability to reduce the need for extensive data annotations, which can be especially difficult to obtain for complex tasks. Sushil et al. [22] investigated the use of LLMs in breast cancer pathology classification, showcasing their potential to alleviate the burden of extensive data labeling. Bansal et al. [23] and Clavié et al. [24] also emphasized the importance of LLMs in improving the generalization of NLP models through innovative annotation techniques. Sun et al. [25] propose clue and reasoning prompting (CARP) for the task of text classification. Gilardi et al. [26] found that ChatGPT outperformed crowd workers in tasks such as relevance, stance, topic, and frame detection, with a zero-shot accuracy approximately 25 percentage points higher than that of human annotators across different datasets. This highlights not only the efficiency of LLMs in text classification but also their ability to handle tasks with high accuracy and reliability.

Lastly, Wang et al. [27] addressed the scalability of text classification by proposing a set of best practices for deploying LLMs to classify large volumes of comments efficiently and cost-effectively. These practices underscore the practical implications of LLMs in real-world settings, where they can offer substantial improvements over traditional methods.

2.3 Node Classification with Graph Neural Networks

Node classification, the task of predicting a category for a node in a network, is an active area of research with various approaches being developed. Among these, semi-supervised learning methods play a significant role in leveraging both labeled and unlabeled data to improve classification accuracy. Abu-El-Haija et al. [28] introduced N-GCN, employing multi-scale graph convolution to capture information from nodes at varying distances in the network. Wang et al. [29] proposed NodeAug, which utilizes data augmentation techniques tailored for graphs to enhance model generalizability and classification performance. Yang et al. [30] developed MGCN, employing contrastive learning to effectively leverage information from labeled and unlabeled nodes, resulting in improved classification accuracy in semi-supervised settings. Xu et al. [31] introduced LC-GNN, emphasizing the importance of label consistency and enlarging the receptive field of nodes to improve classification accuracy.

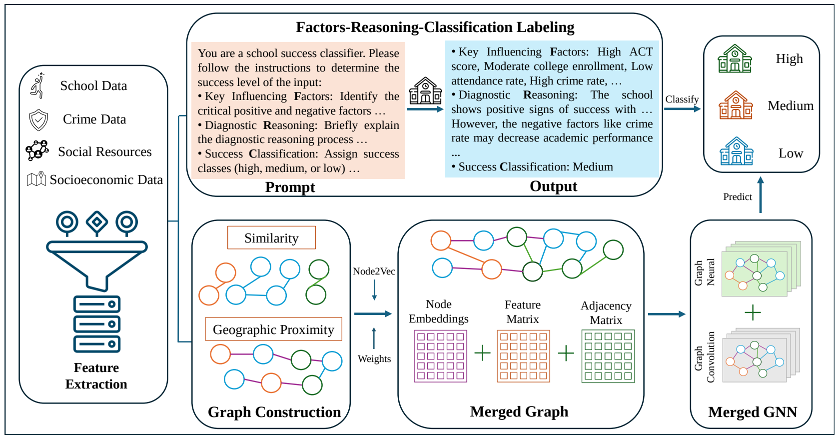

Fig. 1. School Success Prediction with FRC and Merged GNN

In multi-label classification, Xu et al. [32] introduce AdaNN, leveraging gated recurrent units (GRUs) to combine node attributes and network topology at different time points, facilitating adaptation to dynamic graph changes. Zhou et al. [33] propose LANC, employing a label-attentive convolution mechanism to focus on informative neighboring nodes for each label, enhancing classification accuracy in multi-label tasks. Zan et al. Zhang et al. [34] present a novel multi-label relational classifier that identifies similar nodes based on local network structure performs clustering on nodes with known labels.

3 Methodology

In this study, we aim to precisely define and predict the success levels of schools by analyzing a comprehensive set of features. Our methodology is structured around three core components: robust feature extraction, innovative FRC (Factors-Reasoning-Classification) prompting, and strategic Merged Graph integration with Merged GNN, each contributing uniquely to the predictive accuracy of our model. We showed our framework in Figure 1.

3.1 Preprocessing and Feature Extraction

In the feature extraction phase of our study, we utilized a comprehensive approach that integrates multiple datasets, enabling us to effectively capture the intricate dynamics of school environments and their communities. Our dataset included comprehensive details about the spatial distribution of key amenities such as parks and libraries within a 2 km radius of each school, enhancing our analysis with critical insights into the accessibility of important recreational and educational resources. To further enrich our feature set, we grouped schools using the K-means clustering into four different feature categories: school-related metrics, socioeconomic indicators, social amenities, and safety statistics. We determined the optimal number of clusters for each category using the elbow method, enabling precise capture and analysis of the underlying patterns and interrelationships within the data.

Additionally, to mitigate issues with missing data, we implemented an imputation strategy that involves replacing missing values with the mean of available corresponding features for each school. For example, if attainment data were missing for a given year, we calculated the mean from the data of adjacent years to ensure our dataset's continuity and integrity.

3.2 Factors-Reasoning-Classification Labeling

To determine the success levels of schools, we developed a specialized prompting method tailored for Large Language Models (LLMs). This innovative approach fosters a gradual decision-making environment.

The method entails a systematic approach consisting of three steps:

- **Key Influencing Factors Identification:** The model prompts to identify critical factors that influence school success, both positively and negatively. This initial step is instrumental in explaining the multifaceted aspects contributing to a school's overall performance.
- **Diagnostic Reasoning Process:** Model provides a concise explanation of the diagnostic reasoning process, justifying the classification of each school's success level. Model takes the positive and negative influencing factors that might affect school success. This step enables the expression of the thought process behind their classification decisions, enhancing transparency and interpretability.
- **Success Classification:** Based on the identified factors and diagnostic reasoning, the model assigns success classes (high, medium, or low) to each school. This final step synthesizes the input data and user rationale to produce actionable insights regarding school success levels.

This prompting model processes by ingesting the individual features of each school as inputs for detailed analysis. Once the analysis is complete, it assigns a specific class label—high, medium, or low—to each school based on the insights gained. These class labels are compared to human evaluation results and then utilized as the ground truth within our Merged GNN methodology. This integration significantly enhances both the accuracy and reliability of our overall approach.

3.3 Graph Construction

We conceptualize schools as nodes within a network, with their connections representing various interrelationships. This setup forms the basis of a node classification problem, where schools are systematically categorized into three distinct success levels: high, medium, and low.

In the graph construction phase, we define two distinct graph structures, denoted as $G_m = (V, E_m)$, where m specifies the graph (either Similarity Graph (1) or Geographic Proximity Graph (2)), and each graph contains a set of n nodes $V = \{v_1, v_2, \ldots, v_n\}$. The adjacency matrix for each graph is represented as $A_m \in \mathbb{R}^{n\times n}$, and the node feature matrix is indicated by $X \in \mathbb{R}^{n\times F}$, where F represents the number of features for each node. For Similarity Graph, edges are defined based on the cosine similarity between nodes; an edge e^m_{ij} between nodes v_i and v_j is established if their cosine similarity exceeds a predetermined threshold (62%), with the edge weight equal to this similarity value. For Geographic Proximity Graph, edges are determined by the haversine distance between nodes, where an edge is formed if the distance is below another specified threshold (5 km). These thresholds ensure that each node is connected to at least one other node in the graph. The ultimate objective is to use these graph structures to predict success levels, represented by $Y = \{Y_1, Y_2, \ldots, Y_n\}$, where Y_i is labeled as 2 for high, 1 for medium, and 0 for low success.

We employed the Node2Vec algorithm [35] to generate node embeddings that effectively capture the topological proximity of nodes within our graphs. Node2Vec operates by taking a graph $G(V, E)$ as input—where V and E represent nodes and edges respectively—and it learns an n-dimensional embedding for each node. The learned embeddings are then represented as follows:

$$\text{node2vec}(G = (V, E)) \rightarrow \mathbb{R}^{n\times c} \tag{1}$$

where c indicates the number of embedding dimensions per node. These embeddings are integrated with the original node features $X \in \mathbb{R}^{n\times F}$, resulting in an enriched node feature matrix $X \in \mathbb{R}^{n\times(F+c)}$, which combines both the original features and the new embeddings, thus updating the node features for enhanced graph analysis.

3.4 Merged GNN

To leverage the distinct topological information embedded in the two types of graphs representing our school network, we implement a Merged Graph Neural Network (Merged GNN) described in Algorithm 1. This approach combines the node embeddings generated from both graphs using the node2vec algorithm, along with the base node features common to each graph. We called Similarity Graph as G_1 and Geographic Proximity GraphG_2 in the formula. The resultant new node feature matrix is formalized for use in Merged GNN as follows:

$$X_{\text{new}} = \text{Concat}(\text{node2vec}(G_1), \text{node2vec}(G_2), X) \tag{2}$$

We create a combined adjacency matrix by summing the adjacency matrices of the two graphs, each weighted appropriately to optimize the model's performance. We denote the weights for Similarity Graph as γ which is a learnable parameter and for Geographic Proximity Graph as $1-\gamma$. The new adjacency matrix is updated as:

$$A_{\text{new}} = (\gamma \times A_1 + (1-\gamma) \times A_2) \tag{3}$$

In our node classification approach, we employ both Graph Convolutional Network (GCN) [36] and Graph Neural Network (GNN) [37] techniques on the merged graph structure. Given the merged graph G with node feature matrix $X_{\text{new}} \in \mathbb{R}^{n\times(F+2c)}$ and the combined adjacency matrix A_{new}, the GCN operates as follows:

$$H = \widetilde{A}_{\text{new}} X_{\text{new}} \Theta \tag{4}$$

where $\widetilde{A}_{\text{new}}$ is the normalized version of A_{new}, calculated by:

$$\widetilde{A}_{\text{new}} = D^{-\frac{1}{2}} A_{\text{new}} D^{-\frac{1}{2}} + I \tag{5}$$

with D representing the diagonal node degree matrix and I the identity matrix. Each GCN layer i updates as:

$$H^i = \widetilde{A}_{\text{new}} H^{i-1} \Theta \tag{6}$$

and the output from the final layer is computed by:

$$H^{\text{final}} = \sigma(H^l \Theta) \tag{7}$$

Here, σ represents the ReLU activation function, Θ learnable parameter, and l the number of hidden layers.

The GNN component begins by defining the neighborhood N_i for each node v_i, consisting of nodes v_j directly connected to v_i:

$$N_i = \{j : e_{ij} \in E\} \tag{8}$$

Message Passing in GNNs involves transforming node features using a function F, aggregating these transformations from neighbors:

$$F(x_j) = W_j \cdot x_j + b \tag{9}$$

$$\bar{m}_i = G\left(\{W_j \cdot x_j : j \in N_i\}\right) \tag{10}$$

Node features are updated by combining the node's current feature vector with the aggregated messages:

$$h_i = \sigma(K(H(x_i), \bar{m}_i)) \tag{11}$$

where K is a projection neural network (MLP) that integrates the features into a new dimension.

By integrating the outputs from both GCN and GNN layers, we derive the final predictions for our node classification task, effectively capturing both local and global structural information. In Algorithm 1, linear refers to a linear transformation applied to combine features, and softmax is used to normalize the output logits into probabilities.

Algorithm 1 Merged Graph Neural Network (Merged GNN) for School Success Prediction

Require: Graph adjacency matrix A, node features X, number of classes C
Ensure: Predicted class labels for each node
Step 1: Initialize Graph Convolutional Network Layers
Setup transformation matrices $\Theta^{(1)}, \Theta^{(2)}$ for GCN layers
Setup weight matrix W for GNN layer
Step 2: Apply Graph Convolutional Network Layers
for $i = 1$ to 2 **do**
  if $i == 1$ **then**
    $H^{(1)} \leftarrow \sigma(\widetilde{A}X\Theta^{(1)})$ ▷ First GCN layer
  else
    $H^{(2)} \leftarrow \sigma(\widetilde{A}H^{(1)}\Theta^{(2)})$ ▷ Second GCN layer
  end if
end for
Step 3: Apply Graph Neural Network Operations
$H^{(GNN)} \leftarrow \sigma(A(XW))$ ▷ GNN operation using learned weights
Step 4: Integrate GCN and GNN Outputs
$H^{(combined)} \leftarrow H^{(2)} + H^{(GNN)}$
Step 5: Class Prediction
$Z \leftarrow \text{softmax}(\text{Linear}(H^{(combined)}))$ ▷ Compute class probabilities
  return Z ▷ Return the class probabilities

4 Experiments

To assess the performance of our proposed model, FRC, and MergeGraph, we conducted extensive experiments to validate its effectiveness.

4.1 Datasets

We are utilizing various datasets sourced from the City of Chicago and the Chicago Metropolitan Agency for Planning (CMAP) [38] for our research. These datasets include information on crime, parks, libraries, boundaries, demographic, and economic data as shown in Table 1. For the Chicago School Dataset, we analyze essential metrics such as graduation rate, college enrollment percentage, and attendance rate for each school over the years 2014, 2015, and 2016. The Chicago Crime Dataset provides comprehensive crime data from the Chicago Data Portal [39]. We focus on specific crime types—robbery, arson, homicide, and offense—and analyze their correlation with school success. Crime rates around each school within a 2 km radius are aggregated for each distinct year.

Incorporating region-specific public health data [40] enhances our dataset with crucial health statistics related to each school's geographic location. This enrichment includes features such as teen birth rate, poverty level, housing conditions, and per capita income sourced from the Community Data Snapshots (CDS) project. We merge data on Chicago Public Parks [39] and libraries [41] with school data based on latitude and longitude features. Our analysis extends to parks conducive to teenage engagement, focusing on amenities suitable for teenagers. Additionally, we consider the count of libraries near a school, acknowledging their impact on learning opportunities and access to resources like computers, books, and the internet within a 2 km radius.

Table 1. Dataset Description

Node Features	Types
School Dataset	Graduation Rate, Attainment, ACT Scores, College Enrollment, Attendance Rate
Crime Dataset	Robbery, Offense, Arson, Homicide
Socioeconomic Dataset	Race, Age, Employment, and Income Distributions, Community Health Data
Social Resource Dataset	Parks and Libraries

4.2 Baselines

We compared our proposed method with five baselines that can be categorized into two groups: graph-based methods and traditional machine-learning models. To compare traditional machine learning models, we used node embeddings and node features as tabular data.

Graph-based methods:

- GNN [37]: Predicts school success labels by learning node representations based on the underlying graph structure.
- GCN [36]: Classifies school nodes by leveraging their spatial and temporal features along with the features of their neighboring nodes.
- Graph Sage [42]: Aggregates information from neighboring nodes and learns representations based on school features and the collective information of neighboring schools and their features.

Traditional machine learning methods:

- XGBoost [43]: Classifies schools by identifying complex patterns and relationships among the provided features.
- Random Forest [44]: Captures the relationships between input variables and the target by constructing decision trees based on subsets of school features.

4.3 Evaluation Metrics and Experiment Settings

We compared the results obtained with the learned weights in the final adjacency matrix required for our proposed model with the results obtained by trying various weight configurations. Additionally, we conducted a comparative analysis between our model and baseline methods. The performance of our model was assessed using key metrics such as F1-Score and accuracy.

For FRC prompting, we leveraged the Chat-GPT 3.5 Turbo model due to its advanced capabilities, which offered several benefits to our study. With its vast language understanding and generation capabilities, this model enabled more precise predictions. To facilitate manageable input sizes and enhance the computational workflow, we imposed a maximum token limit of 256.

5 Results

5.1 LLM-Based Data Labeling Results

To assess the efficacy of the FRC prompting method, we conducted a comparative analysis against human-labeled data. Three individuals assigned success labels to 50 schools based on the subset of the dataset that is provided to the LLM. We determined the label for each school by selecting the most frequently assigned category. For instance, if two evaluators labeled a school as "medium" while another labeled it as "high", we assigned the label "medium" to that school as human-labeled data and compared it to prompt labels. Figure 2 shows the confusion matrix of labeling. While our prompt model doesn't suggest that a school commonly perceived as "high" by humans is actually "low", only one of the schools identified as "high" by the model is deemed "low" by humans. Regarding "medium" schools, both human evaluation and our model generally yield similar assessments. Notably, there are instances where schools classified as "high" by the model are labeled as "medium" by humans.

We compared human-labels and prompting result to reach F1-score, precision, and recall. Table 2 presents the convergence of our model's scores with those assigned by humans across various evaluation metrics. Remarkably, our model demonstrated comparable performance in all aspects. On average, there's a 65% similarity between human evaluations and prompting results. Consequently, the LLM enabled us to discern all labels without the need for additional human labeling efforts, showcasing its efficiency and accuracy in predictive tasks.

Table 2. FRC and Human Evaluation Comparison

Class	F1-Score	Precision	Recall
High	0.6316	0.5455	0.7500
Medium	0.6522	0.7143	0.6000
Low	0.6857	0.6667	0.7059

5.2 Merged GNN Performance with Different Weights

In our study, we put together two distinct graph structures: one defining the interconnectivity of schools based on node similarity (Similarity Graph), while

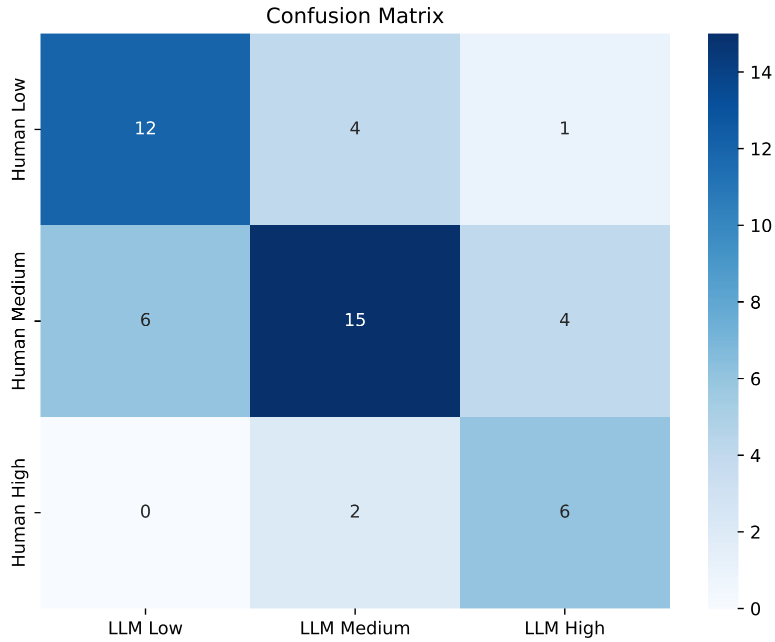

Fig. 2. Confusion Matrix of Human-Labels and LLM-Labels

the other encapsulated the geographical proximity between schools (Geographic Proximity Graph). As detailed in Table 3, we systematically evaluated the impact of different weight distributions on model performance and compared the results of these weights with the weights learned by our model. The Merged GNN assigns weights of 0.3063 and 0.6937 to the Similarity Graph and Geographic Proximity Graph, respectively. The near balance between these weights at (0.3, 0.7) leads to highly competitive results. This distribution outperforms alternative weight assignments. This strategic weighting strategy optimally leverages the unique characteristics of each graph, enhancing the predictive capabilities of our model.

5.3 School Success Prediction Results

As shown in Table 4, our proposed methodology, FRC and Merged GNN, demonstrates notable superiority when compared to existing graph neural network models, XGBoost, and Random Forest classifier. Our proposed model achieves an accuracy of 0.897 and an F1 score of 0.8928, indicating its efficacy in school success prediction task. Comparative analysis reveals the individual performance of graph neural network (GNN), graph convolutional network (GCN), and GraphSAGE on the constituent graphs. Specifically, while GNN achieves better re-

Table 3. Weights For Merge Graph

Weight for G1	Weight for G2	Accuracy	F1 Score	Precision	Recall
0.9	0.1	0.682	0.6711	0.7069	0.6753
0.7	0.3	0.7568	0.7486	0.7858	0.7486
0.5	0.5	0.7872	0.7817	0.8056	0.7906
0.3	0.7	0.897	0.8928	0.8931	0.893
0.1	0.9	0.738	0.7299	0.7578	0.7377
0.3063	**0.6937**	**0.8951**	**0.8914**	**0.892**	**0.8915**

sults for Geographic Proximity Graph (G2) with 0.5297, GCN attains 0.7335 and 0.5925 for Similarity Graph (G1) and Geographic Proximity Graph (G2), respectively. Graph SAGE gives similar results with random forest in terms of Similarity Graph (G1) but fails to learn from Geographic Proximity Graph (G2). Moreover, Random Forest predicts classes in Geographic Proximity Graph (G2) with 70% accuracy, whereas XGBoost's results hover around 0.50 for both graphs. These findings underscore the effectiveness of our proposed model in leveraging the complementary information from merged graphs, offering a robust framework for predictive modeling in complex network data.

Table 4. Comparison

Models	Graphs	Accuracy	F1 Score	Precision	Recall
GNN	G1	0.4053	0.3944	0.3946	0.3952
	G2	0.5297	0.4949	0.5642	0.5148
GCN	G1	0.7335	0.729	0.7545	0.7301
	G2	0.5925	0.5552	0.6266	0.577
Graph SAGE	G1	0.5741	0.4771	0.4887	0.5494
	G2	0.2556	0.1357	0.0852	0.3333
XGBoost	G1	0.4815	0.4618	0.4588	0.4815
	G2	0.5926	0.5956	0.6005	0.5926
Random Forest	G1	0.5556	0.5302	0.5377	0.5556
	G2	0.7037	0.6956	0.7007	0.7037
Merged GNN	**Merged Graph**	**0.897**	**0.8928**	**0.8931**	**0.893**

6 Conclusion

As a consequence, our investigation attempts to forecast school achievement by explaining the multifaceted influences. We introduce a novel methodology by leveraging Factor-Reasoning-Classification (FRC) prompting and Merged GNN integration. The FRC prompting method facilitates the automated identification of school achievement labels through comprehensive knowledge and diagnostic reasoning. Simultaneously, the Merged GNN technique combines two distinct graph structures, one based on school similarity and the other on geographic proximity, to provide a comprehension of school attributes and interconnections. Through extensive experiments on real-world datasets, our methodology shows significant improvements over baseline models. Our approach achieves superior performance in school achievement prediction by effectively integrating multiple graph structures and leveraging advanced predictive modeling techniques.

Acknowledgment

Melike Yildiz Aktas is financially supported by the Turkish Ministry of National Education for her PhD research. Many thanks to Dr. Taoran Ji, Dr. Abdulaziz Alhamadani, and Shailik Sarkar for their invaluable feedback and Priya Nitin Pitre for her help in collecting some datasets used in this study.

References

1. M. Wenger, H. Gärtner, and M. Brunner, “To what extent are characteristics of a school’s student body, instructional quality, school quality, and school achievement interrelated?” *School effectiveness and school improvement*, vol. 31, no. 4, pp. 548–575, 2020.
2. S. M. Daily, M. J. Mann, A. L. Kristjansson, M. L. Smith, and K. J. Zullig, “School climate and academic achievement in middle and high school students,” *Journal of school health*, vol. 89, no. 3, pp. 173–180, 2019.
3. G. Torrats-Espinosa, “Crime and inequality in academic achievement across school districts in the united states,” *Demography*, vol. 57, no. 1, pp. 123–145, 2020.
4. P. W. Githaka, M. W. Gachahi, and J. Mwaruvie, “Social amenities and academic performance in primary schools in gichugu constituency, kenya.” *European Journal of STEM Education*, vol. 4, no. 1, p. 4, 2019.
5. H. Dunga and C. Mafini, “Socio-economic factors influencing the education of the girl child in zomba, malawi,” *International Journal of Social Sciences and Humanity Studies*, vol. 11, no. 2, pp. 20–38, 2019.
6. E. Alyahyan and D. Düştegör, “Predicting academic success in higher education: literature review and best practices,” *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1, p. 3, 2020.
7. A. S. Hashim, W. A. Awadh, and A. K. Hamoud, “Student performance prediction model based on supervised machine learning algorithms,” in *IOP conference series: materials science and engineering*, vol. 928, no. 3. IOP Publishing, 2020, p. 032019.
8. S. McCartney and X. Rosenvasser, “Not your parents’ dorm room: Changes in universities’ residential housing privacy levels and impacts on student success,” *SAGE Open*, vol. 13, no. 2, p. 21582440231178540, 2023.
9. B. Bell, R. Costa, and S. Machin, “Why does education reduce crime?” *Journal of political economy*, vol. 130, no. 3, pp. 732–765, 2022.
10. L. Lochner, “Education and crime,” in *The economics of education.* Elsevier, 2020, pp. 109–117.

11. R. Jin, R. Wu, Y. Xia, and M. Zhao, "What cultural values determine student self-efficacy? an empirical study for 42 countries and economies," *Frontiers in Psychology*, vol. 14, p. 1177415, 2023.
12. N. Souchon, G. Kermarec, D. Trouilloud, and B. Bardin, "Influence of teachers' political orientation and values on their success prediction toward students from different socioeconomic background," *European Review of Applied Psychology*, vol. 70, no. 5, p. 100553, 2020.
13. J. L. Mahoney, R. P. Weissberg, M. T. Greenberg, L. Dusenbury, R. J. Jagers, K. Niemi, M. Schlinger, J. Schlund, T. P. Shriver, K. VanAusdal *et al.*, "Systemic social and emotional learning: Promoting educational success for all preschool to high school students." *American Psychologist*, vol. 76, no. 7, p. 1128, 2021.
14. B. A. Gueldner, L. L. Feuerborn, and K. W. Merrell, *Social and emotional learning in the classroom: Promoting mental health and academic success.* Guilford Publications, 2020.
15. A. Woehr and K. Newman, "Effects of dog ownership and academic success: A correlational study," *Journal of Student Research*, vol. 9, no. 1, 2020.
16. M. N. Bastedo, M. Umbricht, E. Bausch, B.-K. Byun, and Y. Bai, "How well do contextualized admissions measures predict success for low-income students, women, and underrepresented students of color?" *Journal of Diversity in Higher Education*, 2023.
17. C. F. Rodríguez-Hernández, E. Cascallar, and E. Kyndt, "Socio-economic status and academic performance in higher education: A systematic review," *Educational Research Review*, vol. 29, p. 100305, 2020.
18. S. Malik and K. Jothimani, "Enhancing student success prediction with featurex: A fusion voting classifier algorithm with hybrid feature selection," *Education and Information Technologies*, pp. 1–51, 2023.
19. S. Natek and M. Zwilling, "Student data mining solution–knowledge management system related to higher education institutions," *Expert systems with applications*, vol. 41, no. 14, pp. 6400–6407, 2014.
20. A. Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting student performance in higher education institutions using decision tree analysis," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, pp. 26–31, 2018.
21. E. Yukselturk, S. Ozekes, and Y. K. Türel, "Predicting dropout student: an application of data mining methods in an online education program," *European Journal of Open, Distance and e-learning*, vol. 17, no. 1, pp. 118–133, 2014.
22. M. Sushil, T. Zack, D. Mandair, Z. Zheng, A. Wali, Y.-N. Yu, Y. Quan, and A. J. Butte, "A comparative study of zero-shot inference with large language models and supervised modeling in breast cancer pathology classification," 2024.
23. P. Bansal and A. Sharma, "Large language models as annotators: Enhancing generalization of nlp models at minimal cost," *arXiv preprint arXiv:2306.15766*, 2023.
24. B. Clavié and G. Soulié, "Large language models as batteries-included zero-shot esco skills matchers," *arXiv preprint arXiv:2307.03539*, 2023.
25. X. Sun, X. Li, J. Li, F. Wu, S. Guo, T. Zhang, and G. Wang, "Text classification via large language models," *arXiv preprint arXiv:2305.08377*, 2023.
26. F. Gilardi, M. Alizadeh, and M. Kubli, "Chatgpt outperforms crowd workers for text-annotation tasks," *Proceedings of the National Academy of Sciences*, vol. 120, no. 30, Jul. 2023. [Online]. Available: http://dx.doi.org/10.1073/pnas.2305016120
27. R. E. Wang, P. Wirawarn, N. Goodman, and D. Demszky, "Sight: A large annotated dataset on student insights gathered from higher education transcripts," *arXiv preprint arXiv:2306.09343*, 2023.

28. S. Abu-El-Haija, A. Kapoor, B. Perozzi, and J. Lee, "N-gcn: Multi-scale graph convolution for semi-supervised node classification," in *uncertainty in artificial intelligence*. PMLR, 2020, pp. 841–851.
29. Y. Wang, W. Wang, Y. Liang, Y. Cai, J. Liu, and B. Hooi, "Nodeaug: Semi-supervised node classification with data augmentation," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 207–217.
30. X. Yang, Y. Wang, Y. Liu, Y. Wen, L. Meng, S. Zhou, X. Liu, and E. Zhu, "Mixed graph contrastive network for semi-supervised node classification," *ACM Transactions on Knowledge Discovery from Data*, 2022.
31. B. Xu, J. Huang, L. Hou, H. Shen, J. Gao, and X. Cheng, "Label-consistency based graph neural networks for semi-supervised node classification," in *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, 2020, pp. 1897–1900.
32. D. Xu, W. Cheng, D. Luo, Y. Gu, X. Liu, J. Ni, B. Zong, H. Chen, and X. Zhang, "Adaptive neural network for node classification in dynamic networks," in *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2019, pp. 1402–1407.
33. C. Zhou, H. Chen, J. Zhang, Q. Li, D. Hu, and V. S. Sheng, "Multi-label graph node classification with label attentive neighborhood convolution," *Expert Systems with Applications*, vol. 180, p. 115063, 2021.
34. Z. Zhang, H. Wang, L. Liu, and J. Li, "Multi-label relational classification via node and label correlation," *Neurocomputing*, vol. 292, pp. 72–81, 2018.
35. A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 855–864.
36. T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.
37. F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE transactions on neural networks*, vol. 20, no. 1, pp. 61–80, 2008.
38. C. M. A. for Planning. (2023) Cmap data hub. [Online]. Available: https://datahub.cmap.illinois.gov
39. C. P. Department. (2023) Chicago data portal. [Online]. Available: https://data.cityofchicago.org
40. P. C. a. U. Chicago Department of Public Health and Metopio. (2023) Chicago health atlas. [Online]. Available: https://chicagohealthatlas.org/
41. Data.gov. (2023) Libraries - 2022 circulation by location. [Online]. Available: https://catalog.data.gov/dataset/libraries-2022-circulation-by-location
42. J. Liu, G. P. Ong, and X. Chen, "Graphsage-based traffic speed forecasting for segment network with sparse data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 1755–1766, 2020.
43. T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
44. L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.