UNIVERSITY OF ALBERTA

Identifying Early Warning Indicators for High School Dropouts

Surina He, Tarid Wongvorachan, Okan Bulut, Ka Wing Lai

Measurement, Evaluation, and Data Science

University of Alberta, Canada

Content

- Background
- Methods
- Results
- Conclusion

Background

- High school dropouts has been an important issue to be handled at all levels of education.

- Unemployment rate $_{\text{less than high school}}$ > 3 * Unemployment rate $_{\text{bachelor degree or more}}$

- This disparity was especially pronounced during the COVID-19 pandemic, when the average unemployment rate for bachelor's degree holders was approximately 8%, but 17% for high school graduates who had not enrolled in any college (Bureau of Labor Statistics, 2022).

- Therefore, researchers are interested in identifying predictors of high school dropouts to aid in its prevention.

Motivation

- There have been some studies used machine learning algorithms to identify potential predictors, such as random forest (Chung & Lee, 2019; Sara et al., 2015), support vector machine, boosted regression, and post-Lasso (Sansone, 2019).

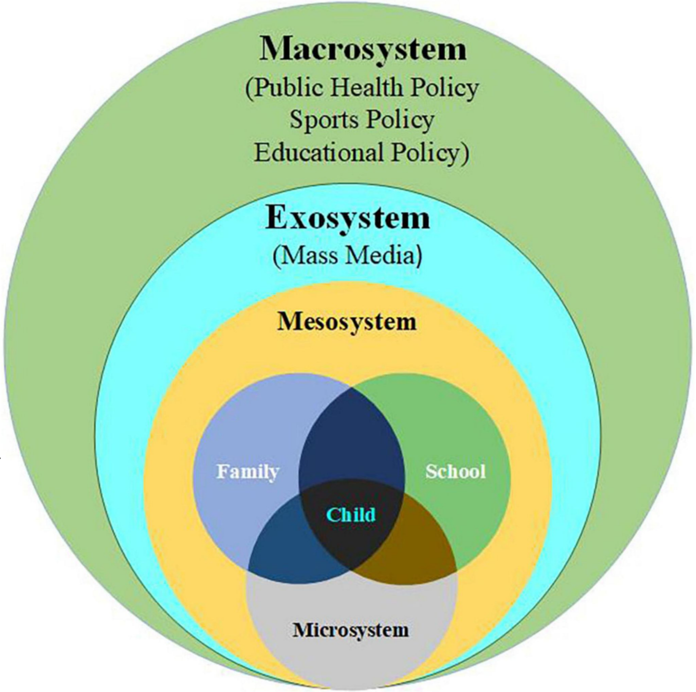
- However, most previous studies:
    a. Examined immutable predictors (i.e., variables that students, teachers, administrators, family and community members have partial or no control over)
    b. Selected predictors without theoretical grounding
    c. Have not tried the deep learning algorithm

- This study aims to use deep learning algorithms to examine mutable/ actionable predictors selected based on Bronfenbrenner's ecological system theory.

Theoretical Framework

- Bronfenbrenner's ecological system theory:
  a. Microsystem: directly interact with student
  b. Mesosystem: interaction between microsystem
  c. Exosystem: indirectly impact student
  d. Macrosystem: societal and cultural contexts

- Factors from individuals, microsystems, and mesosystem have greatest impact on individual development.



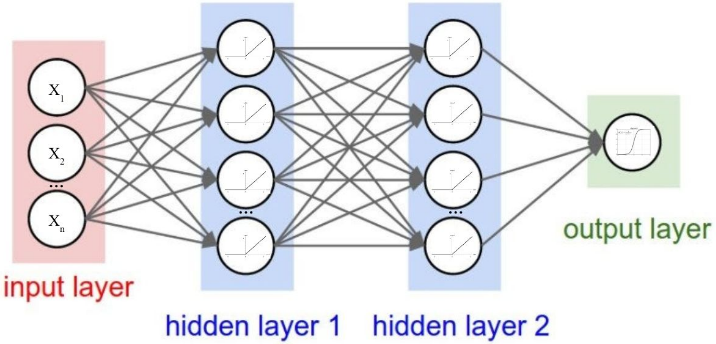
Method

- **Data**: High School Longitudinal Study of 2009 (HSLS:09)

  (1st round in 2009, 9th grade → 2nd round in 2012, 12th grade → 3rd round in 2016)
- **Sample**: 23,503 9th graders from 944 schools, and their parents, teachers, school principals
- **Data pre-processing**:
  a. Removing cases that parental questionnaire was not answered by their biological/stepparents
  b. Removing variables with a missing rate of 30% or higher
  c. Imputing missing values using Random Forest imputation
  d. Removing predictors that were not correlate with high school dropouts (1=dropout, 0= not dropout)
  e. Balancing the samples in two dropout group using Synthetic Minority Oversampling Technique-Nominal
- **Final data**: 38 predictors, 22,612 (11,306 for each class) 9th graders.

Method

- **Deep learning algorithm:**
  - Keras package in Python to build a deep neural network model via Tensorflow.
  - Four layers:
    1. Input layer
    2. 1st hidden layer: 128 nodes, Rectified Linear Unit (ReLU) activation function
    3. 2nd hidden layer: 32 nodes, ReLu function
    4. Output layers: 1 node, the sigmoid function



Results

- Feature importance:
    a. GPA
    b. Teacher expectations

Predictors	Weight (Feature Importance)
The 9th grade GPA	0.1381 ± 0.0044
Teacher expectations	0.0550 ± 0.0077
Family socioeconomic status	0.0480 ± 0.0059
Counselor expectations	0.0452 ± 0.0058
Mothers' educational level	0.0428 ± 0.0041
Principal expectations	0.0415 ± 0.0052
Fathers' educational level	0.0379 ± 0.0044
Students' sense of school belonging	0.0355 ± 0.0043
Students' college enrollment status	0.0355 ± 0.0037
School climate	0.0331 ± 0.0058

Results

- **GPA:**

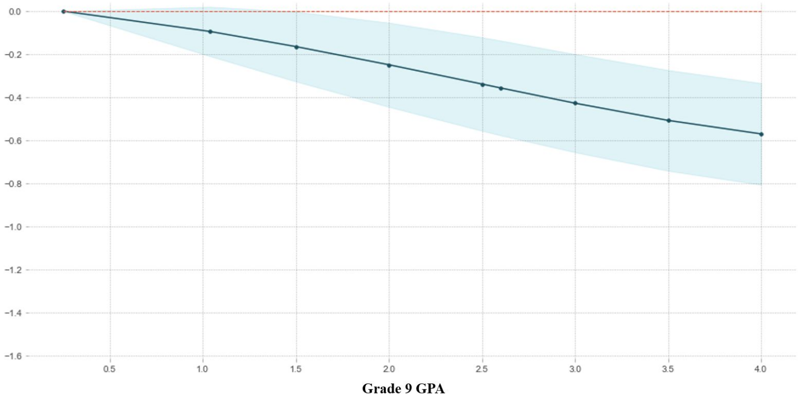
  The higher GPA, the less dropout

- **Teacher expectation:**

  A quadratic relationship between expectation and dropout

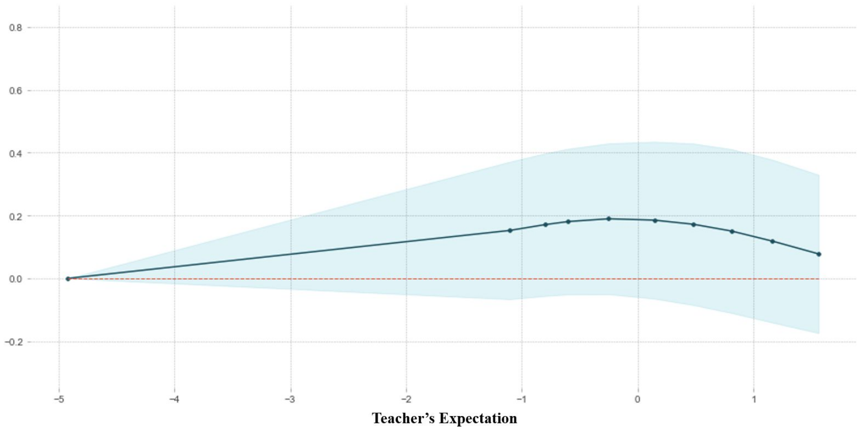
Partial Dependence Plot for Grade 9 GPA

Number of unique grid points: 9



Partial Dependence Plot for Teacher’s Expectation

Number of unique grid points: 10



Discussion

- Using a deep learning system, the current study identified the top ten predictors of dropouts, seven of which are actional predictors.

- GPA and future-oriented goals (i.e., teacher expectation, counselor expectation, principal expectations) are two most important predictors.
    - High GPA could protect students from dropping out of high school, and moderate level expectations from others are the most beneficial.

- There are many studies examined the role of educational expectations to students, which is a good news to educational practitioners.
    - We can encourage students to commit themselves to meaningful long-term educational goals, seek to benefit from their educational experiences, which would lead them to monitor their progress toward their goals.

Discussion

- The quadratic relationship between the expectation and high school dropouts indicates that the role of expectation is a mixed blessing. This quadratic pattern was also found in the relationship between expectation and students' academic achievement (He et al., 2022).

- The potential moderators could be control and autonomy support, which were framed in both self-determination theory (Joussemet et al., 2008; Moe et al., 2020; Schiffrin et al., 2019) and the Beliefs, Expectations, Autonomy Support, and Relationships (BEAR) model (Froiland, 2021).

Discussion

- **Limitations**
    a. It will be better, if we compared the performance of deep learning with other ML algorithms
    b. We manually selected original 67 predictors from 1000+ predictors in the data pool based on they ecological system theory, which may introduce potential biases in the results.
- **Contributions**
    a. This is the first attempt to use deep learning algorithms to identify potential predictors.
    b. This study added evidence to the importance of educational expectations, as well as its' relationship with high school dropouts.

Thank you for your listening !

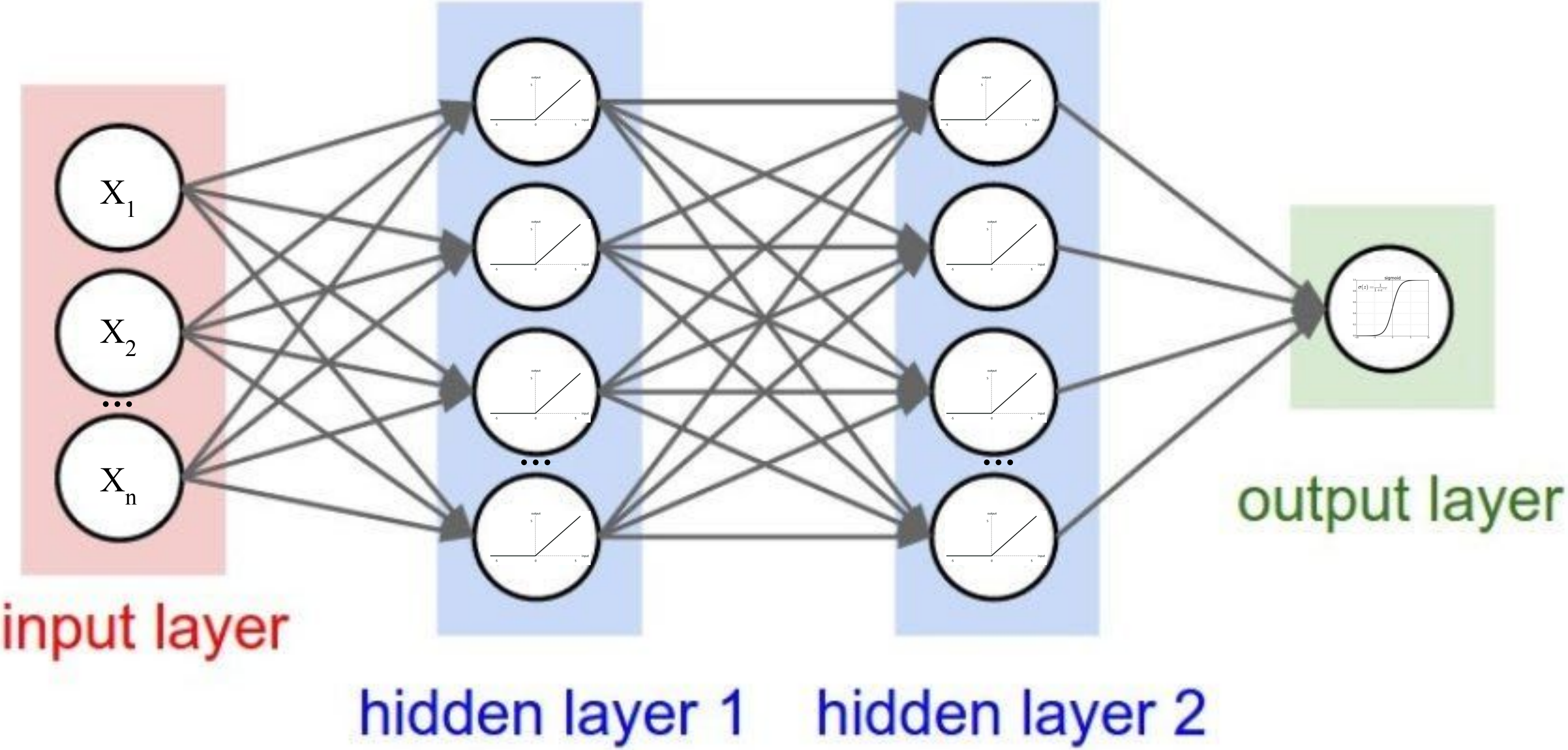
For questions and comments, please contact:

Surina He surina1@ualberta.ca

Tarid Wongvorachan wongvora@ualberta.ca

Okan Bulut bulut@ualberta.ca

Ka Wing Lai kawing2@ualberta.ca



X1
X2
Xn
input layer
hidden layer 1
hidden layer 2
output layer