

Predicting Cognitive Distortions in Youth Using Advanced Machine Learning Algorithms

Masoumeh. Abdollahi¹, Najmeh. Naseri², Fatemeh. Hayati^{3*}, Ayoub. Ghasemi⁴

¹ Ph.D. in Educational Psychology, Bafg.C., Islamic Azad University, Yazd, Iran

² M.A. in Clinical Psychology, Zar.C., Islamic Azad University, Kerman, Iran

³ M.A. in Educational Management, Shahid Bahonar University of Kerman, Kerman, Iran



⁴ M.A., Department of Psychology, Faculty of Social Sciences, Razi University, Kermanshah, Iran

* Corresponding author email address: f.hayati67@yahoo.com

Editor

John S. Carlson
Distinguished Professor of the
Department of Educational
Psychology, Michigan State
University, East Lansing, MI,
United
carlsoj@msu.edu

Reviewers

Reviewer 1: Parvaneh Mohammadkhani
Professor, Department of Clinical Psychology, University of Rehabilitation Sciences
and Social Health, Tehran, Iran. Email: Pa.mohammadkhani@uswr.ac.ir
Reviewer 2: Mahdi Khanjani
Associate Professor, Department of Psychology, Allameh Tabataba'i University,
Tehran, Iran.
Email: khanjani_m@atu.ac.ir

1. Round 1

1.1. Reviewer 1

Reviewer:

In the discussion of university students as a vulnerable population, the manuscript cites multiple stressors but does not explicitly justify why undergraduate students (as opposed to adolescents or graduate students) are optimal for predictive modeling. A brief comparative justification is recommended.

The paragraph beginning “Cognitive distortions do not develop in isolation...” introduces attachment and childhood trauma, yet these constructs are not included in the predictive model. Please clarify whether they were excluded due to data limitations or theoretical scope.

The statement “most existing research relies on traditional statistical methods” is accurate but broad. Consider adding 1–2 concrete examples (e.g., linear regression, SEM) to strengthen the methodological critique.

The study aim is clearly stated; however, it would benefit from explicitly naming the main benchmark model (e.g., multiple linear regression) to clarify the comparative framework used in the analyses.

The description “ensemble-based methods and nonlinear learners” is conceptually vague. Please list the algorithms explicitly at first mention (e.g., Random Forest, XGBoost) before using categorical labels.

Table 1 reports descriptive statistics but does not indicate sample size (N). Including N in the table caption or header would improve clarity and reporting standards.

The statement “all-or-nothing thinking and emotional reasoning showed relatively higher mean values” is descriptive but not contextualized. Consider briefly comparing these values to normative or prior student samples.

Authors uploaded the revised manuscript.

1.2. Reviewer 2

Reviewer:

The use of “stratified convenience sampling” requires clarification. Please specify the stratification variables (e.g., gender, academic field) and explain how stratification was operationalized within a convenience framework.

The claim “The final sample size was determined to be sufficient for machine learning modeling” should be supported by either a rule-of-thumb (e.g., observations-to-features ratio) or a citation to ML methodological literature.

While the cognitive distortion instrument is described as “standardized and psychometrically validated,” its exact name, original developer, and example item should be briefly reported for transparency and replicability.

Internal consistency coefficients are mentioned but not reported. Please include Cronbach’s alpha (or alternative reliability indices) for all major scales in the current sample.

The manuscript states that missing data were handled using “appropriate imputation techniques,” but does not specify which method (e.g., mean imputation, k-NN, multiple imputation). This omission limits reproducibility.

Authors uploaded the revised manuscript.

2. Revised

Editor’s decision after revisions: Accepted.

Editor in Chief’s decision: Accepted.