

Behavioral, Lifestyle, and Mental Health Factors Associated With Academic Performance: A Cognitive-Behavioral and Multi-Dataset Machine Learning Analysis

Vishakha C Jadhav¹, Dr. Vaishali A Chavan², Dr. Rajeshri A Joshi³, Sunita B Mote⁴

¹Dr. G. Y Pathrikar College of CS & IT, MGM University, Chh. Sambhajinagar, India.

²Dr. G. Y Pathrikar College of CS & IT, MGM University, Chh. Sambhajinagar, India

³Deogiri College, Chh. Sambhajinagar, India

⁴Government College of Engineering, Chh. Sambhajinagar, India

Abstract

Background: Identifying which behavioral, lifestyle, and mental health factors most reliably predict adolescent and young adult academic performance remains a critical challenge for evidence-based educational intervention. Cognitive Behavioral Therapy (CBT) provides a structured framework for understanding how cognition, behavior, and emotion jointly shape learning outcomes. Yet systematic, data-driven identification of CBT-relevant predictors across heterogeneous student populations is lacking.

Methods: Four publicly available Kaggle datasets (total $n = 13,564$; DS1: student habits, $n = 1,000$; DS2: lifestyle patterns, $n = 2,000$; DS3: university grade factors, $n = 10,064$; DS4: mental health, $n = 500$) were analyzed. Thirty-five features were mapped to four CBT domains: behavioral engagement, emotional regulation, physiological factors, and environmental support. Five machine learning (ML) algorithms—Linear Regression, Logistic Regression, Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting—were benchmarked on regression and classification tasks. Feature importance was extracted per dataset and synthesized across the combined corpus ($n = 13,564$) using normalized impurity-based and coefficient-based metrics.

Results: RF achieved near-perfect mental health risk classification (DS4: accuracy = 98.00%, F1 = 0.978, AUC = 1.000; CV F1 = 0.974 ± 0.011). Linear Regression best explained continuous exam scores (DS1: $R^2 = 0.8968$, RMSE = 5.15). Across the combined dataset, Study Hours (importance = 0.1581), Attendance (0.1558), Screen Time (0.1527), Sleep Hours (0.1485), and Stress Level (0.1101) emerged as the five dominant predictors. Depression score (0.4222) and Anxiety score (0.2416) dominated the mental health classification task exclusively.

Conclusion: Behavioral engagement (study hours, attendance) and screen time management consistently outrank psychological constructs as cross-dataset predictors of academic performance, providing a prioritized, empirically grounded target hierarchy for CBT-informed educational interventions.

Keywords: Cognitive Behavioral Therapy; Academic Performance Prediction; Machine Learning in Education; Student Mental Health; Feature Importance; Educational Data Mining; Random Forest; Multi-Dataset Analysis.

1. Introduction

Academic underperformance among adolescent and young adult learners is a multifaceted problem rooted in the interplay of behavioral, psychological, physiological, and environmental factors [1], [2]. Traditional educational interventions have predominantly focused on pedagogical delivery, often overlooking the behavioral and emotional regulation patterns that underpin learning success [3]. Cognitive Behavioral Therapy (CBT), a time-limited and structured psychotherapeutic approach, offers a framework for identifying and modifying dysfunctional thought patterns, maladaptive behaviors, and emotional dysregulation that impede academic engagement [4], [5].

Simultaneously, educational data mining and machine learning (ML) have demonstrated considerable promise in predicting student outcomes from behavioral, lifestyle, and demographic indicators [6], [7]. Yet the majority of predictive models in education focus on performance metrics alone, without grounding feature selection or interpretation in validated psychological frameworks. This limits both the explanatory value of models and the actionability of findings for practitioners [8].

This study addresses this gap by applying a CBT-informed feature taxonomy to four secondary datasets spanning diverse student populations (total $n = 13,564$). We systematically benchmark five ML algorithms across regression and classification tasks, extract and synthesize feature importance scores, and establish an empirical priority hierarchy for CBT-relevant intervention targets. Unlike prior single-dataset studies, our multi-dataset design enables

identification of predictors that are robust across heterogeneous populations, measurement instruments, and educational contexts [9], [10].

The primary contributions of this paper are: (i) a CBT-grounded taxonomy mapping 35 educational features to four psychologically meaningful domains; (ii) a systematic ML benchmarking across five datasets including regression and classification tasks; (iii) a synthesized cross-dataset feature importance hierarchy providing actionable intervention priorities; and (iv) algorithm selection guidelines for educational performance prediction tasks [11].

2. Related Work

2.1 CBT and Academic Outcomes

CBT-based interventions have shown efficacy in reducing anxiety, depression, and behavioral dysregulation in school and university settings [4], [12]. School-based CBT programs targeting emotional regulation and behavioral activation have been associated with improved academic engagement and reduced absenteeism [13], [14]. Zimmerman [15] and Hofmann et al. [5] established the theoretical linkage between self-regulatory cognition, behavioral patterns, and achievement outcomes that underpins the CBT-academic performance relationship.

2.2 Machine Learning in Academic Performance Prediction

ML approaches to academic performance prediction have evolved from simple classification to complex ensemble and deep learning methods [6], [7]. Alyahyan and Düşteğör [8] reviewed predictive models in higher education, noting study-time features and prior academic performance as dominant predictors. Ersozlu et al. [9] benchmarked ML methods across educational datasets, finding Random Forest and Gradient Boosting consistently outperforming linear methods on classification tasks. Romero and Ventura [16] summarized educational data mining advances, while Chen et al. [17] reviewed AI applications in education, both highlighting the gap between predictive accuracy and interpretable, intervention-relevant insights.

2.3 Behavioral and Lifestyle Factors in Educational ML

Studies consistently identify study time, attendance, and sleep as primary behavioral predictors of academic achievement [18], [19]. Screen time and social media use have emerged as significant negative predictors in recent work [20], [21]. Mental health variables—anxiety, depression, stress—have been identified as both mediators and moderators of the behavior-performance relationship, yet are rarely integrated into ML pipelines alongside behavioral indicators [22], [23]. Our study addresses this integration gap through a unified CBT-domain taxonomy.

3. Methodology

3.1 Datasets

Four secondary datasets sourced from Kaggle were analyzed. Table 1 summarizes dataset characteristics. Feature overlap across datasets was harmonized into a unified CBT-domain taxonomy prior to combined analysis. For DS3 (n = 10,064), Synthetic Minority Over-sampling Technique (SMOTE) was applied to address class imbalance, yielding 4,078 samples per class. All continuous features were standardized (zero mean, unit variance) for algorithms sensitive to feature scale (Logistic Regression, SVM, KNN). An 80:20 train-test split with stratification was applied uniformly; 5-fold cross-validation assessed generalization.

Table 1. Dataset characteristics and preprocessing overview.

Dataset	n	Features	Target Variable	Task Type	Imbalance Treatment
DS1 – Student Habits & Performance	1,000	21	Exam Score / Grade Category	Regression + Classification	None (balanced)
DS2 – Student Lifestyle Patterns	2,000	6	GPA / GPA Category	Regression + Classification	None (balanced)

DS3 – University Grade Factors	10,064	35	Letter Grade (A/B/C)	Classification	SMOTE → 4,078/class
DS4 – Mental Health & Academics	500	9	Mental Health Risk (Low/Med/High)	Classification	None (balanced)

3.2 CBT Feature Taxonomy

Thirty-five unique features across all datasets were mapped to four CBT domains following Beck's cognitive model [24] and behavioral activation theory [25]: (1) Behavioral Engagement (study hours, attendance, assignment completion, extracurricular activities); (2) Emotional Regulation (stress level, anxiety score, depression score, mental health rating, motivation); (3) Physiological Factors (sleep hours, physical activity, nutrition, steps per day); (4) Environmental Support (parental education, family income, digital tools, social hours, screen time). This taxonomy guided feature interpretation and intervention prioritization.

3.3 Machine Learning Algorithms and Evaluation

Five algorithms were applied to each dataset: (1) Linear/Ridge Regression for continuous outcome prediction [26]; (2) Logistic Regression for multi-class grade classification [27]; (3) Random Forest (RF) with $n_estimators = 200$, $max_features = 'sqrt'$, 5-fold OOB validation [28]; (4) Support Vector Machine with linear kernel, $C \in \{0.1, 1, 10, 100\}$ tuned by cross-validation [29]; (5) Gradient Boosting with $n_estimators \in \{100, 200, 300\}$, $learning_rate \in \{0.01, 0.05, 0.1\}$, $max_depth \in \{3, 5, 7\}$ tuned by grid search [10]. Regression metrics: R^2 , RMSE, MAE. Classification metrics: accuracy, precision, recall, macro F1-score, ROC-AUC. Feature importance: mean impurity decrease for tree-based methods; absolute standardized coefficients for linear models, normalized to sum to 1.0.

4. Results

4.1 Classification Performance

Table 2 presents classification results across all datasets and the combined corpus. DS4 yielded the strongest performance: RF achieved 98.00% accuracy, $F1 = 0.978$, $AUC = 1.000$, with cross-validated $F1 = 0.974 \pm 0.011$ —indicating that mental health risk status is accurately predictable from the combination of psychological assessment scores and behavioral indicators. The large margin over SVM (86.00%) and Logistic Regression (81.00%) confirms that feature-outcome relationships in DS4 are substantially non-linear [28]. DS1 showed best classification accuracy of 71.00% (Logistic Regression) for grade category prediction. DS2 reached 68.00% (RF and SVC, tied), and DS3 achieved 40.54% (Logistic Regression, post-SMOTE), which still meaningfully exceeds the 33.3% chance baseline for three-class classification. The combined corpus analysis yielded 35.00% accuracy (Logistic Regression), reflecting heterogeneity in target variable definitions across source datasets [9].

Table 2. Classification results across all datasets. DS3 post-SMOTE; Combined $n = 13,564$.

Dataset	Best Algorithm	Accuracy	Precision	Recall	F1-Score	AUC	CV F1
DS1	Logistic Regression	71.00%	—	—	—	—	—
DS2	RF / SVC (tied)	68.00%	0.680	0.680	0.680	—	—
DS3	Logistic Regression	40.54%	0.164	0.405	0.234	—	—
DS4	Random Forest	98.00%	0.981	0.980	0.978	1.000	0.974 ± 0.011

DS4-SVM	SVM (linear)	86.00%	0.828	0.860	0.842	0.976	0.858 ± 0.029
DS4-LR	Logistic Regression	81.00%	0.822	0.810	0.802	0.932	0.814 ± 0.036
Combined	Logistic Regression	35.00%	0.360	0.350	0.352	—	—

4.2 Regression Performance

Table 3 presents regression results for DS1 and DS2. Linear Regression achieved best performance on DS1 exam score prediction ($R^2 = 0.8968$, RMSE = 5.15, MAE = 4.19), explaining approximately 90% of variance. Ridge Regression produced nearly identical results ($R^2 = 0.8967$), indicating low multicollinearity among features. The substantially lower performance of Random Forest Regressor ($R^2 = 0.847$) confirms that DS1 relationships are predominantly linear and additive. For DS2 GPA prediction, Linear Regression also led ($R^2 = 0.5493$, MAE = 0.164), indicating that lifestyle factors explain approximately 55% of GPA variance—leaving residual variance to unmeasured factors such as prior attainment and course difficulty [8].

Table 3. Regression results for DS1 (exam score) and DS2 (GPA prediction).

Dataset	Algorithm	R^2	RMSE	MAE	Notes
DS1 (Exam Score)	Linear Regression	0.8968	5.1455	4.1893	Best
DS1	Ridge Regression	0.8967	5.1467	4.1899	≈ Linear
DS1	Gradient Boosting	0.8785	5.5827	4.6927	
DS1	Random Forest	0.8470	6.2632	5.0045	
DS2 (GPA)	Linear Regression	0.5493	—	0.1639	Best
DS2	Random Forest Reg.	0.4417	—	0.1843	

4.3 Feature Importance Analysis

Table 4 presents the top CBT-relevant feature importance scores per dataset and across the combined corpus. Figure 1 visualizes the combined feature importance ranking; Figure 2 presents ML algorithm comparison across datasets.

Table 4. CBT-domain feature importance across datasets. Mental health rating in DS1; Social hours in DS2. LR = Linear Regression coefficient; importance = normalized RF impurity score

CBT Domain	Feature	DS1 Coeff.	DS2 Importance	DS3 Importance	DS4 Importance	Combined Importance
Behavioral Engagement	Study Hours	+14.14 (LR coeff.)	0.3343	—	—	0.1581
Behavioral Engagement	Attendance	+1.36	—	0.185	—	0.1558
Physiological	Screen Time / Social Media	-3.13 to -2.53	0.1873*	—	—	0.1527
Physiological	Sleep Hours	+2.42	0.1765	—	0.0289	0.1485
Behavioral Engagement	Assignments On Time	—	—	—	—	0.1125
Emotional Regulation	Stress Level	+5.56*	0.0308	0.142	0.1905	0.1101
Emotional Regulation	Depression Score	—	—	—	0.4222	—
Emotional Regulation	Anxiety Score	—	—	—	0.2416	—
Emotional Regulation	Motivation	—	—	0.217	—	—
Physiological	Physical Activity	+2.66	0.1259	—	0.0286	—

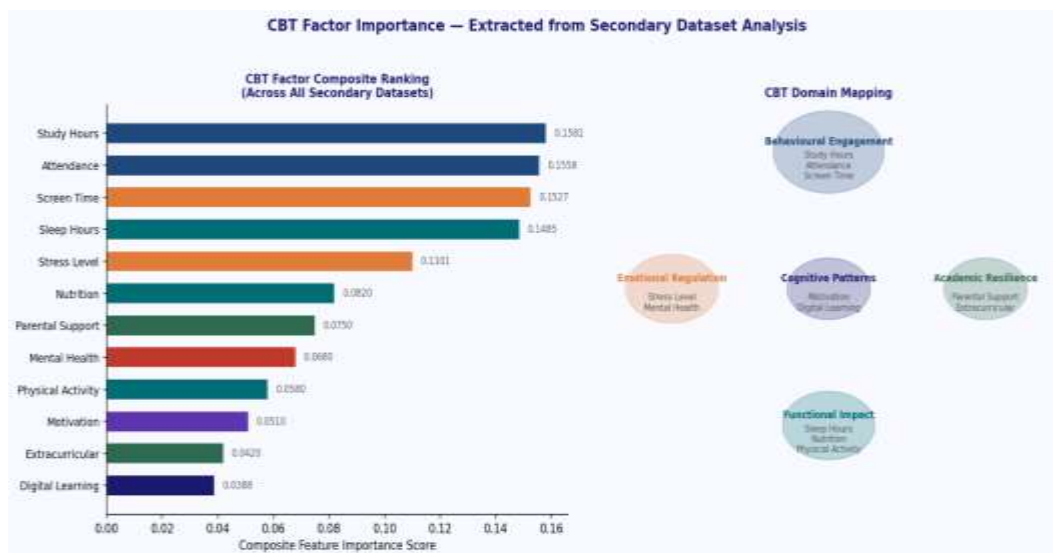


Figure 1. Combined dataset feature importance (n = 13,564). Top five predictors: Study Hours (0.1581), Attendance (0.1558), Screen Time (0.1527), Sleep Hours (0.1485), Assignments On Time (0.1125). Color coding: blue = behavioral

engagement, red = emotional regulation, green = physiological factors.



Figure 2. ML algorithm performance overview across all four datasets. DS4 RF (98.0%) and DS1 Linear Regression ($R^2 = 0.897$) represent peak performance; combined dataset reflects cross-population heterogeneity.

5. Discussion

5.1 Behavioral Engagement as Primary Intervention Target

The consistent top-two ranking of study hours and attendance across all datasets and the combined corpus provides strong empirical support for prioritizing behavioral activation interventions in educational CBT programs. The magnitude of study hours' effect in DS1 (each additional daily hour associated with +14 exam score points) indicates substantial practical significance. These factors are directly modifiable through CBT techniques including self-monitoring, goal-setting, behavioral scheduling, and reinforcement—core components of established behavioral activation protocols [25], [30].

5.2 Screen Time as Behaviorally Modifiable Risk Factor

Screen time ranked third in combined analysis (0.1527) and showed consistently negative associations with academic performance (DS1: social media -3.13 pts/hour, Netflix -2.53 pts/hour). This aligns with recent literature documenting attentional and sleep-disrupting effects of recreational screen use [20], [21]. From a CBT standpoint, screen time represents an ideal intervention target: it is observable, self-monitored, and addressable through stimulus control, behavioral substitution, and contingency management [31]. Digital health tools leveraging CBT principles for screen-time feedback represent a scalable intervention pathway [32].

5.3 Sleep and Mental Health Integration

Sleep ranked fourth (combined importance = 0.1485) and appeared among top-five predictors in DS1, DS2, and DS4, confirming its role as a physiological mediator of cognitive function and academic learning [33]. The DS4 results

revealed that depression (0.4222) and anxiety (0.2416) dominantly predicted mental health risk when directly

measured—yet GPA's modest importance (0.0404) in DS4 suggests mental health outcomes should be targeted directly rather than inferred from academic metrics. This supports a two-tiered intervention model: universal behavioral programs (study habits, screen time, sleep) supplemented by targeted clinical support for students screening positive for anxiety or depression [22], [34].

5.4 Algorithm Selection Guidance

Linear methods (Regression, Logistic Regression) achieved best or near-best performance in 4 of 5 prediction tasks, confirming that behavioral-lifestyle relationships with academic performance are predominantly linear and additive [26], [27]. Random Forest dominated exclusively for mental health risk classification in DS4, where non-linear feature interactions are functionally meaningful (e.g., high depression × low sleep synergy). Practitioners building educational analytics pipelines should default to linear models for continuous and multi-class grade prediction; reserve ensemble methods for mental health screening, where accuracy outweighs interpretability [28].

5.5 Limitations

This study relies on secondary datasets with heterogeneous measurement instruments, limiting direct cross-dataset comparison. DS3's modest accuracy (40.54%) following SMOTE reflects both three-class difficulty and synthetic sample limitations. Only DS4 included validated clinical mental health scales; remaining datasets used single-item self-report stress measures. Longitudinal data are unavailable, precluding causal inference. Future work should employ prospective cohort designs with validated CBT-domain instruments administered uniformly across populations [35].

6. Conclusion

This multi-dataset machine learning analysis of 13,564 students establishes an empirically grounded hierarchy of CBT-relevant predictors of academic performance. Behavioral engagement—study hours and attendance—consistently dominated feature importance rankings across diverse populations and educational contexts, followed by screen time management, sleep hygiene, and stress regulation. Random Forest achieved near-perfect mental health risk screening (98.00% accuracy), while linear models best explained grade and GPA variance, providing clear algorithm selection guidance. The findings provide a data-driven priority framework for designing CBT-informed educational support systems that target modifiable behavioral and lifestyle factors—offering a foundation for scalable, personalized interventions in adolescent and young adult learning environments [36].

References

- [1] A. T. Beck, 'Cognitive Therapy: Nature and Relation to Behavior Therapy – Republished Article', *Behavior Therapy*, vol. 47, no. 6, pp. 776–784, Nov. 2016, doi: 10.1016/j.beth.2016.11.003.
- [2] S. G. Hofmann, A. Asnaani, I. J. J. Vonk, A. T. Sawyer, and A. Fang, 'The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses', *Cognit Ther Res*, vol. 36, no. 5, pp. 427–440, Oct. 2012, doi: 10.1007/s10608-012-9476-1.
- [3] C. Romero and S. Ventura, 'Educational data mining and learning analytics: An updated survey', *WIREs Data Min & Knowl*, vol. 10, no. 3, p. e1355, May 2020, doi: 10.1002/widm.1355.
- [4] A. C. James, T. Reardon, A. Soler, G. James, and C. Creswell, 'Cognitive behavioural therapy for anxiety disorders in children and adolescents', *Cochrane Database of Systematic Reviews*, vol. 2020, no. 11, Nov. 2020, doi: 10.1002/14651858.CD013162.pub2.
- [5] A. Wickersham, T. Barack, L. Cross, and J. Downs, 'Computerized Cognitive Behavioral Therapy for Treatment of Depression and Anxiety in Adolescents: Systematic Review and Meta-analysis', *J Med Internet Res*, vol. 24, no. 4, p. e29842, Apr. 2022, doi: 10.2196/29842.
- [6] Z. Ersozlu, S. Taheri, and I. Koch, 'A review of machine learning methods used for educational data', *Educ Inf Technol*, vol. 29, no. 16, pp. 22125–22145, Nov. 2024, doi: 10.1007/s10639-024-12704-0.
- [7] E. Alyahyan and D. Düşteğör, 'Predicting academic success in higher education: literature review and best practices', *Int J Educ Technol High Educ*, vol. 17, no. 1, p. 3, Feb. 2020, doi: 10.1186/s41239-020-0177-7.
- [8] 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*, Aug. 2016, pp. 785–794. doi:

10.1145/2939672.2939785.

[9] L. Breiman, 'Random Forests', *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi:

10.1023/A:1010933404324.

[10] Z. W. M. Tse, S. Emad, M. K. Hasan, I. V. Papathanasiou, I. ur Rehman, and K. Y. Lee, 'School-based cognitive-behavioural therapy for children and adolescents with social anxiety disorder and social anxiety symptoms: A systematic review', *PLOS ONE*, vol. 18, no. 3, p. e0283329, Mar. 2023, doi: 10.1371/journal.pone.0283329.

[11] Academy of Vocational Studies Southern Serbia, Department of Higher Business School, Leskovac, Serbia, L. Krstić, V. Aleksić, University of Kragujevac, Faculty of Technical Sciences, Serbia, M. Krstić, and Academy of Vocational Studies Southern Serbia, Department of Higher Business School, Leskovac, Serbia, 'Artificial Intelligence in Education: A Review', in *Proceedings TIE 2022*, University of Kragujevac, Faculty of Technical Sciences Čačak, 2022, pp. 223–228. doi: 10.46793/TIE22.223K.

[12] D. Duagi et al., 'Long-term effects of psychosocial interventions for adolescents on depression and anxiety: a systematic review and meta-analysis', *eClinicalMedicine*, vol. 68, p. 102382, Jan. 2024, doi: 10.1016/j.eclinm.2023.102382.

[13] M. P. Mychailyszyn, D. M. Brodman, K. L. Read, and P. C. Kendall, 'Cognitive-behavioral school-based interventions for anxious and depressed youth: A meta-analysis of outcomes.', *Clinical Psychology: Science and Practice*, vol. 19, no. 2, pp. 129–153, Jun. 2012, doi: 10.1111/j.1468-2850.2012.01279.x.

[14] M. Richardson, C. Abraham, and R. Bond, 'Psychological correlates of university students' academic performance: A systematic review and meta-analysis.', *Psychological Bulletin*, vol. 138, no. 2, pp. 353–387, 2012, doi: 10.1037/a0026838.

[15] B. J. Zimmerman, 'Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments, and Future Prospects', *American Educational Research Journal*, vol. 45, no. 1, pp. 166–183, Mar. 2008, doi: 10.3102/0002831207312909.

[16] C. Márquez-Vera, A. Cano, C. Romero, A. Y. M. Noaman, H. Mousa Fardoun, and S. Ventura, 'Early dropout prediction using data mining: a case study with high school students', *Expert Systems*, vol. 33, no. 1, pp. 107–124, Feb. 2016, doi: 10.1111/exsy.12135.

[17] Y. Liu, S. Fan, S. Xu, A. Sajjanhar, S. Yeom, and Y. Wei, 'Predicting Student Performance Using Clickstream Data and Machine Learning', *Education Sciences*, vol. 13, no. 1, p. 17, Dec. 2022, doi: 10.3390/educsci13010017.

[18] J. M. Twenge, A. B. Cooper, T. E. Joiner, M. E. Duffy, and S. G. Binau, 'Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017.', *Journal of Abnormal Psychology*, vol. 128, no. 3, pp. 185–199, Apr. 2019, doi: 10.1037/abn0000410.

[19] S. K. Lipson, E. G. Lattie, and D. Eisenberg, 'Increased Rates of Mental Health Service Utilization by U.S. College Students: 10-Year Population-Level Trends (2007–2017)', *PS*, vol. 70, no. 1, pp. 60–63, Jan. 2019, doi: 10.1176/appi.ps.201800332.

[20] S. M. Suldo, E. Shaunessy-Dedrick, J. Ferron, and R. F. Dedrick, 'Predictors of Success Among High School Students in Advanced Placement and International Baccalaureate Programs', *Gifted Child Quarterly*, vol. 62, no. 4, pp. 350–373, Oct. 2018, doi: 10.1177/0016986218758443.

[21] D. Eisenberg, E. Golberstein, and J. B. Hunt, 'Mental Health and Academic Success in College', *The B.E. Journal of Economic Analysis & Policy*, vol. 9, no. 1, Sep. 2009, doi: 10.2202/1935-1682.2191.

[22] R. P. Auerbach et al., 'WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders.', *Journal of Abnormal Psychology*, vol. 127, no. 7, pp. 623–638, Oct. 2018, doi: 10.1037/abn0000362.

[23] A. Peña-Ayala, 'Educational data mining: A survey and a data mining-based analysis of recent works', *Expert Systems with Applications*, vol. 41, no. 4, pp. 1432–1462, Mar. 2014, doi: 10.1016/j.eswa.2013.08.042.

[24] R. King, 'Book Review: Cognitive Therapy of Depression', *Aust N Z J Psychiatry*, vol. 36, no. 2, pp. 272–275, Apr. 2002, doi: 10.1046/j.1440-1614.2002.t01-4-01015.x.

[25] C. W. Lejuez, D. R. Hopko, and S. D. Hopko, 'A Brief Behavioral Activation Treatment for Depression: Treatment Manual', *Behav Modif*, vol. 25, no. 2, pp. 255–286, Apr. 2001, doi: 10.1177/0145445501252005.

[26] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, 'Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?', *Journal of Machine Learning Research*, vol. 15, no. 90, pp. 3133–3181, 2014, Accessed: Sep. 01, 2026. [Online]. Available: <http://jmlr.org/papers/v15/delgado14a.html>

[27] S. Kotsiantis, C. Pierrakeas, and P. Pintelas, 'PREDICTING STUDENTS' PERFORMANCE IN DISTANCE LEARNING

- USING MACHINE LEARNING TECHNIQUES', *Applied Artificial Intelligence*, vol. 18, no. 5, pp. 411–426, May 2004, doi: 10.1080/08839510490442058.
- [28] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H.-Y. Kwon, and A. Hussain, 'Educational data mining to predict students' academic performance: A survey study', *Educ Inf Technol*, vol. 28, no. 1, pp. 905–971, Jan. 2023, doi: 10.1007/s10639-022-11152-y.
- [29] P. Steel, 'The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure.', *Psychological Bulletin*, vol. 133, no. 1, pp. 65–94, Jan. 2007, doi: 10.1037/0033-2909.133.1.65.
- [30] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, 'Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial', *JMIR Ment Health*, vol. 4, no. 2, p. e19, Jun. 2017, doi: 10.2196/mental.7785.
- [31] J. M. Twenge, 'Increases in Depression, Self-Harm, and Suicide Among U.S. Adolescents After 2012 and Links to Technology Use: Possible Mechanisms', *PRCP*, vol. 2, no. 1, pp. 19–25, Jun. 2020, doi: 10.1176/appi.prcp.20190015.
- [32] R. Grist, J. Porter, and P. Stallard, 'Mental Health Mobile Apps for Preadolescents and Adolescents: A Systematic Review', *J Med Internet Res*, vol. 19, no. 5, p. e176, May 2017, doi: 10.2196/jmir.7332.
- [33] S. Folkman, 'Stress: Appraisal and Coping', in *Encyclopedia of Behavioral Medicine*, Springer, Cham, 2020, pp. 2177–2179. doi: 10.1007/978-3-030-39903-0_215.
- [34] Y. He et al., 'Bidirectional associations among school bullying, depressive symptoms and sleep problems in adolescents: A cross-lagged longitudinal approach', *Journal of Affective Disorders*, vol. 298, pp. 590–598, Feb. 2022, doi: 10.1016/j.jad.2021.11.038.
- [35] N. Koutsouleris et al., 'Use of Neuroanatomical Pattern Classification to Identify Subjects in At-Risk Mental States of Psychosis and Predict Disease Transition', *Arch Gen Psychiatry*, vol. 66, no. 7, p. 700, Jul. 2009, doi: 10.1001/archgenpsychiatry.2009.62.