Westborough High School Mental Health Predictive Analysis

An Interactive Qualifying Project
Submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE
In Partial Fulfilment of the Requirements for the
Degree of Bachelor of Science

By
Aukkawut Ammartayakun
Ben Kresge
Brian Borghesani
Khoi Pham
Lindsey Mraz

Submitted to
Prof. Benjamin Nephew
Prof. Dmitry Korkin
Prof. Angela Incollingo Rodriguez
Prof. Richard Lopez

April 25, 2024

Abstract

Over the years, mental health challenges, particularly those of students, have increased, so the need to develop targeted diagnostic and treatment tools has increased. In view of substantial research that highlights the massive variation in symptomatic disorders in mental health, the effort to develop an intervention strategy can benefit from personal insight. The study studies the predictive capacity of random forest classification (RFC) to predict mental health risk scores in mental health data from Westborough High School. Our results show differences in mental health outcomes among different student groups and highlight the need for personalized interventions rather than a one-size-fits-all approach.

Keywords: Random Forest, Mental Health Prediction, High School, Middle School

Contents

Chapter 1

Introduction

1.1 Introduction

Mental health remains a major challenge, especially in light of the rise of COVID-19, which affects many people, including adolescents (de Figueiredo et al., 2021; Marques de Miranda et al., 2020). Within the US alone, among high-school adolescents during 2019 and 2021, the risk of suicide planning increased from 19.9% to 23.6%, besides the consideration of the suicide attempts that jump from 24.1% to 30% (Gaylor et al., 2023). Lipson et al. (2022) also highlights in their study of more than 300,000 students that the number of students who met at least one of the criteria of mental health problems increased drastically to nearly 50% during the span of 2013 to 2021. All of these facts provide evidence that the mental health issue of adolescents is no longer a small problem.

The assessment can be originally made using the psychometric test (Chorpita et al., 2000; Kroenke et al., 2001; Spitzer et al., 2006). The increasing trend of machine learning in the field of education and mental health (Kučak et al., 2018; Pongpaichet et al., 2022; Thieme et al., 2020) elucidates the solution for determining the risk group. Extensive research (Lo, 2023; Tate et al., 2020; Usher and Curran, 2017) has been conducted to determine and predict students at risk and provide early intervention. However, most of them have the same suggestion that machine learning prediction could not be used in the clinical setting.

This study is an extension of the previous work by Lo (2023) that proposed the integration of the random forest model with the academic and psychometric characteristics collected from grade 9 students at Westborough High School. This work extensively expands the students' cohort from only grade 9 to include grades 7, 8, 9, and 11.

The significance of this work lies in the diathesis-stress model (Ingram and Luxton, 2005) where the interplay between the predisposition and the environmental factor are the causes of mental disorders. In this case, the latter are captured from the psychometric test and the academic information which can help indicate the risk group among the rest of the students based on the environmental factor.

1.2 Problem Statement

The lack of a reliable system to accurately identify mental health risks in middle- and high-school students in Westborough hinders early intervention efforts. Additionally, there is limited understanding of the specific academic and psychometric features that are most strongly associated with these risks, making it difficult to focus on prevention and support strategies.

In this study, risk prediction will be made in middle and high school students in Westborough. Let the features (both academic and psychometric) that correspond to the ith student be $\mathbf{x}_i$; we aim to find the random forest classifier f such that $f(\mathbf{x}_i)$ is within the tolerable error range ε, able to predict the risk of those particular students. We then learn from that classifier f by looking at the importance of features and the characteristics of the data to learn and identify those features that will help early intervention.

1.3 Objective

The primary objective of this study is to develop and validate a robust machine learning model, specifically utilizing a random forest classifier, to accurately predict mental health risks among middle and high school students in Westborough. By leveraging academic and psychometric data, the model aims to identify at-risk students with a high degree of accuracy within an acceptable error margin. Through this approach, the study seeks to enhance understanding of the key features that correlate with mental health risks, thereby facilitating targeted early intervention strategies.

Chapter 2

Background and Literature Review

2.1 Psychopathology of Common Adolescents Mental Disorders

While there are many mental disorder that commonly occurs in adolescents (Castellanos-Ryan et al., 2016; Maddux and Winstead, 2020), in this study, the depression and anxiety disorders will be discussed as those are clinically diagnosable (Chorpita et al., 2000; Kroenke et al., 2001; Spitzer et al., 2006) with the psychometric tests used in this study.

2.1.1 Diathesis-Stress Model of Mental Disorder

The diathesis-stress model is a psychological theory that explains the onset of mental disorders through the interaction between a pre-existing vulnerability (diathesis) and environmental stressors (Ingram and Luxton, 2005). This model is particularly useful in understanding how certain adolescents develop mental disorders such as depression and anxiety while others do not, despite facing similar life challenges.

In the context of adolescents, the diathesis can include biological factors such as genetics or neurobiological anomalies as well as psychological factors like early life experiences, personality traits, or previous trauma (Ingram and Luxton, 2005). Environmental stressors relevant to adolescents often include academic pressures, peer relationships, family dynamics, and significant life changes or adverse events.

According to this model, the presence of a diathesis does not guarantee the development of a mental disorder; rather, it makes an individual more susceptible to developing a disorder when combined with sufficient environmental stress (Ingram and Luxton, 2005).

2.1.2 Depression Disorder

Depression in adolescents is a serious and prevalent mental health disorder characterized by persistent sadness, loss of interest in activities, and a significant decrease in overall daily functioning. Clinically, adolescent depression is often diagnosed through the use of psychometric tests such as the Patient Health Questionnaire (PHQ-9), which evaluates the severity of depressive symptoms over the past two weeks (Kroenke et al., 2001). This disorder can significantly impact various aspects of a young person's life, including academic performance, social interactions, and physical health. Factors contributing to depression in this age group include biological, psychological, and environmental stressors. The symptoms are varied but commonly include feelings of worthlessness, changes in sleep and appetite, difficulty concentrating, and recurrent thoughts of death or suicide (Maddux and Winstead, 2020).

2.1.3 Anxiety Disorder

Anxiety disorders in adolescents encompass a range of conditions, including generalized anxiety disorder (GAD), panic disorder, and social anxiety disorder, among others (Maddux and Winstead, 2020). These disorders are characterized by excessive and persistent worry that is difficult to control, causing significant distress or impairment in social, academic, or other important areas of functioning. Tools like the Generalized Anxiety Disorder 7-item (GAD-7) scale and the Revised Child Anxiety and Depression Scale (RCADS) are commonly used to assess the presence and severity of anxiety symptoms (Chorpita et al., 2000; Spitzer et al., 2006). Symptoms typically include restlessness, fatigue, difficulty concentrating, irritability, muscle tension, and sleep disturbances. Early identification and intervention are crucial as these disorders are comorbid with other mental disorders like depression (Maddux and Winstead, 2020).

2.2 Psychometric Tests

Psychometric tests are standardized tools used to measure psychological variables such as symptoms, traits, and cognitive functions (Eun and Boram, 2019). In the context of assessing mental health in adolescents, specific tests have been adopted to evaluate symptoms of depression and anxiety. These tests are essential for diagnosing mental health conditions and are used to gather data that inform the feature set of our predictive model.

2.2.1 PHQ-9

The Patient Health Questionnaire-9 (PHQ-9) is a widely utilized clinical instrument for diagnosing depression and assessing its severity (Eun and Boram, 2019; Kroenke et al., 2001). This test consists of nine items, which correspond directly to the criteria for depression as specified in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Respondents are asked to rate how often they have been bothered by problems such as "little interest or pleasure in doing things" or "feeling down, depressed, or hopeless" over the past two weeks. Each item is scored on a scale from 0 ("not at all") to 3 ("nearly every day"), allowing for a maximum score of 27 (Kroenke et al., 2001). Scores of 10 or higher typically suggest moderate to severe depression, indicating a need for further evaluation and possibly clinical intervention.

2.2.2 GAD-7

The Generalized Anxiety Disorder 7-item (GAD-7) scale is a questionnaire used to screen for and assess the severity of generalized anxiety disorder (Spitzer et al., 2006). This tool evaluates the frequency of symptoms such as excessive worry, restlessness, and irritability over the last two weeks. Like the PHQ-9, responses range from 0 ("not at all") to 3 ("nearly every day"). A cumulative score of 10 or more is indicative of possible generalized anxiety disorder, prompting further investigation and potential treatment. The GAD-7 is recognized for its reliability and validity and is commonly used alongside the PHQ-9 in various clinical settings (Eun and Boram, 2019; Spitzer et al., 2006).

2.2.3 RCADs

The Revised Child Anxiety and Depression Scale (RCADs) is designed to assess symptoms of anxiety and depression in children and adolescents. This tool comprises several subscales that evaluate different types of anxiety disorders and depressive disorders, reflecting the symptomatology as defined by the DSM-5. Each item on the RCADs asks respondents to rate how often they experience specific symptoms on a scale from 0 ("never") to 3 ("always"). The comprehensive nature of this scale makes it particularly useful for detecting a range of emotional difficulties in younger populations (Chorpita et al., 2000). The RCADs can provide detailed insights into the specific types of anxiety or depressive disorders an adolescent may be experiencing, which can guide tailored intervention strategies. In general, T scores of more than 65 are considered borderline clinical and need further evaluation (Chorpita et al., 2000).

2.3 Random Forest

Random forest is an effective ensemble learning algorithm within the supervised machine learning domain. Its core strength lies in the strategic aggregation of multiple decision trees to achieve more robust classification and regression results (Biau, 2012; Breiman, 1996, 2001; James et al., 2023).

The process begins with bootstrapping (Efron, 1979), where each decision tree within the forest is trained on a unique subset of the original training dataset created through random sampling with replacement. This ensures diversity among trees within the ensemble.

Crucially, random forest combats the issue of high correlation between trees, which is common in bagging models (Biau, 2012; Breiman, 1996), by injecting randomness into the tree building process. At each node split during tree construction, the algorithm does not consider the full set of features. Instead, it randomly selects a subset of features. The optimal split is then determined solely from this constrained feature subset, forcing the diversity and reducing overfitting tendencies (Biau, 2012; Breiman, 2001; James et al., 2023). In the sense of classification, the results of each learner are aggregated using the majority vote.

2.4 Datasets

This study utilizes anonymized data from middle and high school students in Westborough, Massachusetts, focusing on academic and psychometric features across four different grade levels—7, 8, 9, and 11. The dataset is comprised of:

- Grade 7 students: 256 samples, including 120 females and 136 males.
- Grade 8 students: 263 samples, with 128 females and 135 males.
- Grade 9 students: 251 samples, consisting of 138 females and 113 males.
- Grade 11 students: 242 samples, with 133 females and 109 males.

In total, the dataset includes 1,012 data points. For each data point, the dataset distinguishes between middle school and high school students, recording features relevant to their respective educational levels.

2.4.1 Middle School Features

Middle school students (grades 7 and 8) are administered the RCADs test. There are the following features in the dataset

- Gender: The student's gender (male or female).
- Grade: The grade level of the student (7, or 8).
- Race / Ethnicity: The student's race or ethnic background.
- ELL (English Language Learner): Indicates whether the student was not born in the U.S. and whose native language is other than English.
- SPED (Special Education): Indicates whether the student is receiving special education services.
- 504: Indicates whether the student has a 504 plan, which provides accommodations based on disability.
- Core Class Avg T1: Average score in core classes for Term 1.
- Core Class Avg T2: Average score in core classes for Term 2.

- Absence: Number of times the student was absent.
- Tardy: Number of times the student was tardy.
- Dismissal: Number of times the student was dismissed early.
- I think about death.: Responses to the RCADs question about thoughts of death.
- Total Depression Score: Cumulative score from depression-related questions in RCADs.
- Depression T Score: Standardized RCADs score reflecting depression severity compared to the population of students who are in a similar age range.
- Total Anxiety Score: Cumulative score from anxiety-related questions in RCADs.
- Anxiety T Score: Standardized RCADs score reflecting anxiety severity compared to the population of students who are in a similar age range.
- Total Anxiety and Depression: Combined score of anxiety and depression assessments.
- Total T Score: Standardized RCADs score of the total anxiety and depression compared to the population of students who are in a similar age range.

2.4.2 High School Features

Middle school students (grades 9 and 11) are administered the PHQ-9 and GAD-7 test. There are the following features in the dataset

- Gender: The student's gender (male or female).
- Grade: The grade level of the student (9 or 11).
- Race / Ethnicity: The student's race or ethnic background.
- ELL (English Language Learner): Indicates whether the student was not born in the U.S. and whose native language is other than English.

- SPED (Special Education): Indicates whether the student is receiving special education services.
- 504: Indicates whether the student has a 504 plan.
- Avg Q1: Average score for Quarter 1.
- Avg Q2: Average score for Quarter 2.
- Avg Level: Average academic complexity level.
- Absence: Number of times the student was absent.
- Tardy: Number of times the student was tardy.
- Dismissal: Number of times the student was dismissed early.
- Thoughts that you would be better off dead or of hurting yourself in some way?: Responses to the survey question about self-harm thoughts in PHQ-9 (Endorse question 9)
- GAD-7 Total: Total score from the GAD-7.
- GAD-7 Rating: Rating based on the GAD-7 total score.
- PHQ-9 Total: Total score from the PHQ-9.
- PHQ-9 Rating: Rating based on the PHQ-9 total score.

Chapter 3

Methodology

3.1 Exploratory Data Analysis

The Exploratory Data Analysis (EDA) in this study primarily focuses on the application of two-way Analysis of Variance (ANOVA) to investigate the effects of gender and race/ethnicity on mental health outcomes, specifically depression and anxiety risk scores among middle and high school students. Moreover, the feasibility of using random forests is explored through clustering in *t*-distributed Stochastic Neighbor Embedding (*t*-SNE).

3.1.1 Data Preparation

Prior to conducting the ANOVA, the data was cleaned and preprocessed to ensure that it was suitable for analysis. This included handling missing values and formatting the columns into a suitable name for processing. The dataset was then segmented based on school level (middle school and high school) to conduct separate analyses for each group. In this stage, descriptive statistics are calculated.

3.1.2 Variable Selection

The dependent variables for the ANOVA were the mental health risk scores derived from the psychometric tests (Depression T Score, Anxiety T Score, and specific questionnaire items like thoughts about death for middle school; GAD-7 and PHQ-9 scores for high school). The independent variables were gender and race/ethnicity. These were treated as fixed factors in the model, with each category (e.g., male, female; White, Asian, etc.) analyzed to assess its impact on the dependent variables.

3.1.3 Model Specification

A two-way ANOVA model was specified for each dependent variable to explore the main effects (ME) of gender and race/ethnicity as well as the interaction effects (IE) between these two factors. This model helps in understanding whether the differences in mental health scores are attributable to one of the factors independently or a combination of both.

3.1.4 *t*-distributed Stochastic Neighbor Embedding (*t*-SNE)

The data are dimensionality reduced via the *t*-distributed Stochastic Neighbor Embedding (*t*-SNE) algorithm to facilitate visualization in a two-dimensional space. Subsequently, each data point was labeled according to its categorization into either the risk or no-risk group as defined by the thresholds established in the Background section of this study. The optimal separation between clusters of risk and no-risk groups would suggest that the given covariates are capable of enabling effective classification. The degree of cluster distinction serves as a preliminary indicator of the potential efficacy of these covariates in achieving high-performance risk classification.

3.2 Association and Causality Assessment

In this work, the focus is on discerning the association between various covariates and comorbidity features with mental health outcomes in adolescents. To achieve this, feature importance metrics are derived from the random forest model. These metrics allow the identification of features that have the most significant impact on the model's predictions, thereby suggesting potential associations with mental health risks.

Feature importance is calculated based on how much each feature contributes to the accuracy of the model across the many decision trees within the random forest. For each stratification (on grade, gender, race, and combination of gender and race), new random forests are trained on a subset of the dataset. Then, the feature importance is evaluated.

The model is evaluated using a 10-fold cross-validation, and the random forest's task is to perform a binary classification of the risk and no-risk group students using the threshold defined in the Background section.

Chapter 4

Results

4.1 ANOVA test

A two-way ANOVA was conducted to assess the impact of gender and race/ethnicity on the mental health risk scores of middle and high school students. The analysis considered main effects (ME) and interaction effects (IE) between these two factors.

4.1.1 Middle School

Depression T Scores

Significant main effects were observed for gender (F=20.21, p<0.001), indicating a substantial difference in depression scores between genders, with one gender exhibiting higher scores than the other. A significant main effect was also found for race/ethnicity (F=7.05, p<0.001). However, the interaction effect between gender and race/ethnicity was not significant (F=0.31, p=0.731), suggesting that the combined effect of these variables does not significantly influence depression scores.

Anxiety T Scores

There were significant main effects for gender (F=35.74, p<0.001) and race/ethnicity (F=4.82, p<0.01). The interaction effect was not significant (F=2.02, p=0.133), indicating that the influence of gender on anxiety scores is consistent across racial/ethnic groups.

Questions About Death

Neither gender (F=2.42, p=0.12) nor race/ethnicity (F=0.29, p=0.745) showed significant differences, and their interaction was also non-significant (F=0.09, p=0.915).

4.1.2 High School

GAD-7 Scores

There were significant differences observed in GAD-7 scores across both race/ethnicity (F=4.9, p=0.03) and gender (F=26.0, p<0.01), with higher anxiety levels reported by White students compared to Asian students and by females compared to males. The interaction between gender and race was not significant (F=0.1, p=0.7).

PHQ-9 Scores

Significant differences were found across genders (F=17.7, p<0.001), with females reporting higher depression scores than males. No significant effects were noted for race (F=1.1, p=0.3) or for the interaction between gender and race (F=0.6, p=0.44).

Endorse Question 9

No significant differences were detected across gender (F=1.6, p=0.21), race/ethnicity (F=0.0, p=0.94), or their interaction (F=0.0, p=0.98).

4.2 *t*-Distributed Stochastic Neighbor Embedding (*t*-SNE)

In this phase of the study, students were categorized based on their risk levels as either no/low risk or high risk, utilizing the RCADS scoring criteria detailed in section 2.2.3. This categorization was applied to assess students' mental health status in grades 7 and 8. The results of the *t*-Distributed Stochastic Neighbor Embedding (t-SNE) analysis are depicted in Figure 4.1 for grade 7 and Figure 4.2 for grade 8. These figures illustrate the clusters formed by t-SNE on the covariates, effectively differentiating between no-risk and high-risk students. The visualization through t-SNE provides a clear demarcation, indicating distinct groupings based on the risk levels as determined by the RCADS scores. This separation suggests that the covariates used hold significant predictive power and are potentially useful for identifying students at risk. This also implies strong evidence that the machine learning model for predicting the risk groups based on the covariates used here is feasible and should perform well. A similar conclusion can be drawn from the *t*-SNE plot of high school students, where PHQ-9 and GAD-7 define risk groups. However, for high school students, the clusters are more spread, and some points are mixed between groups, indicating that it is harder to differentiate between two groups with only the covariates that we have.

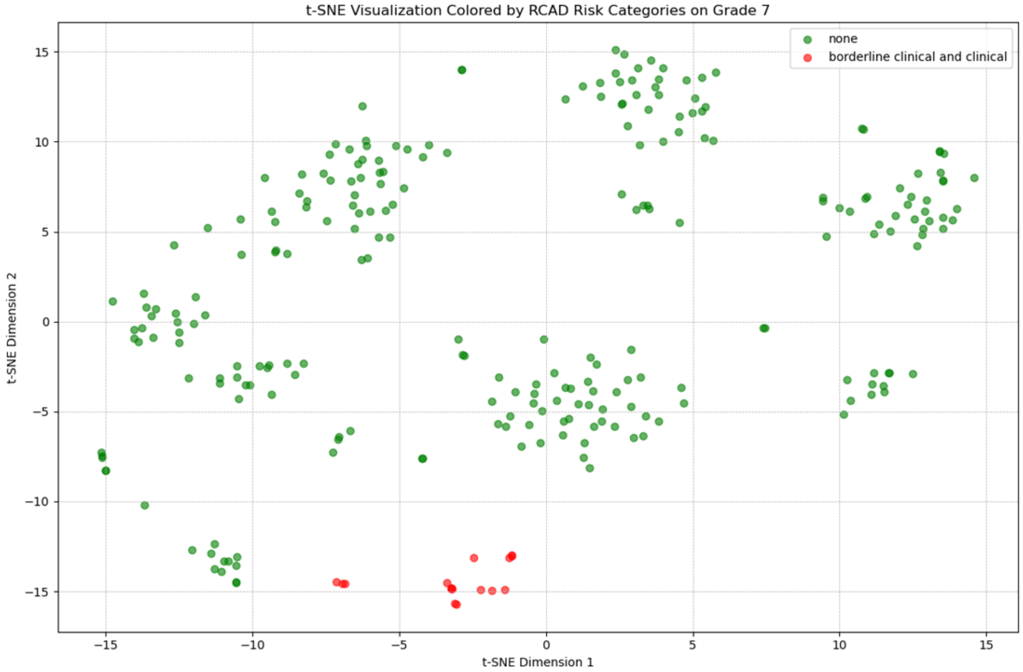

Figure 4.1: t-SNE representation of the risk group defined by RCADs within the grade 7 students ($n = 256$). This representation shows the cluster that differentiates between no-risk ($n = 242$, green clusters) and high-risk students ($n = 14$, red cluster at the bottom).

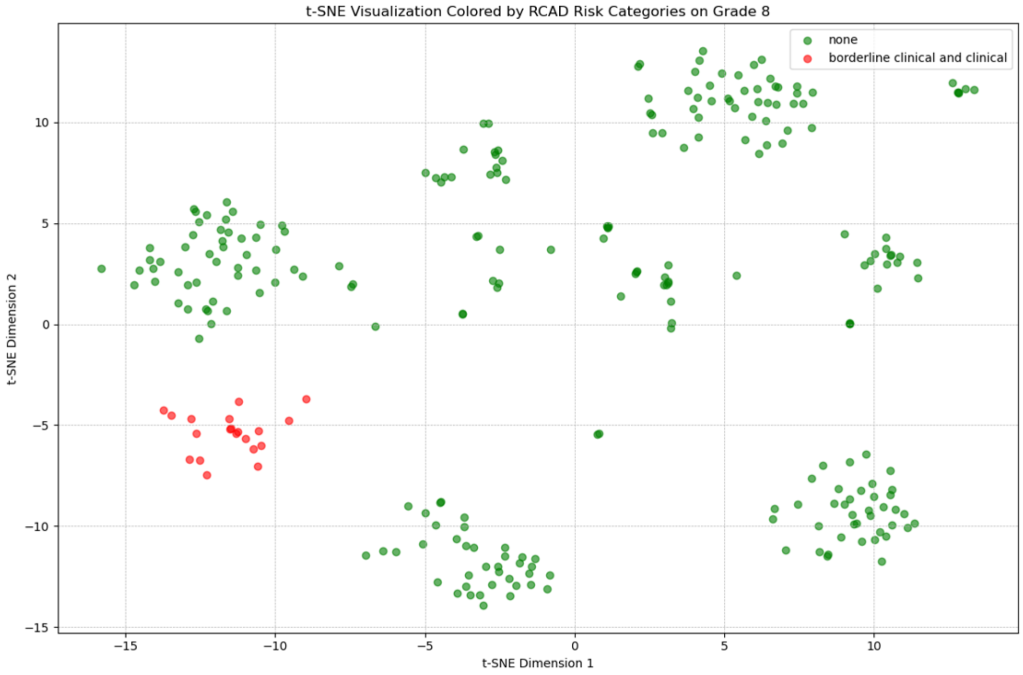

Figure 4.2: t-SNE representation of the risk group defined by RCADs within the grade 8 students ($n = 263$). This representation shows the cluster that differentiates between no-risk ($n = 243$, green clusters) and high-risk students ($n = 20$, red cluster around the left side).

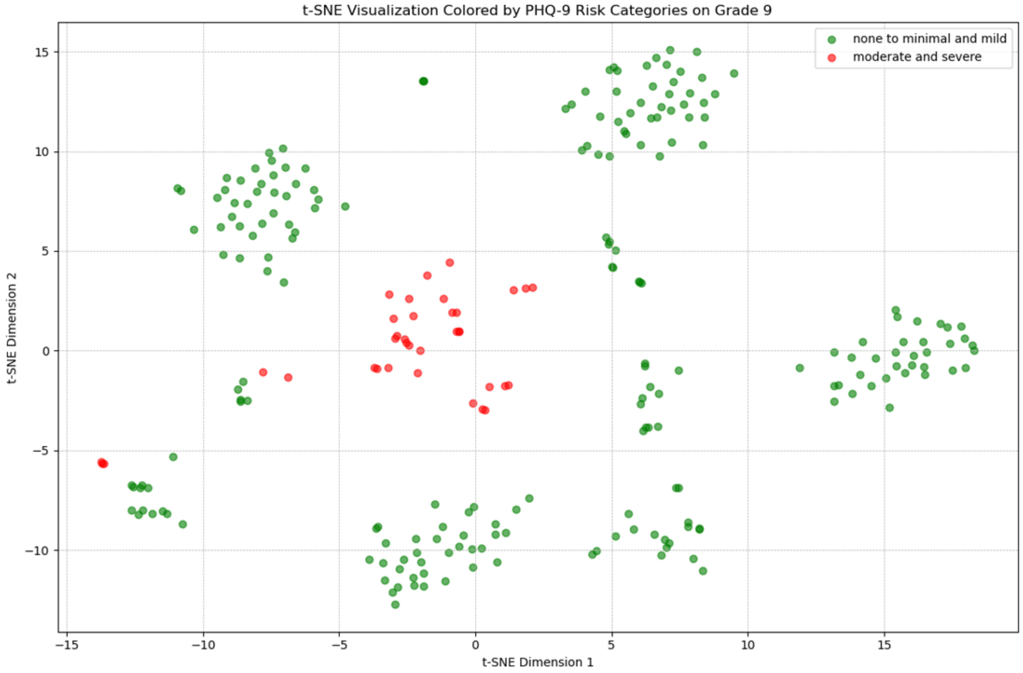

Figure 4.3: t-SNE representation of the risk group defined by PHQ-9 within the grade 9 students ($n = 251$). This representation shows the cluster that differentiates, although not perfectly, between no-risk ($n = 217$, green clusters) and high-risk students ($n = 34$, red cluster in the middle with outliers on the left).

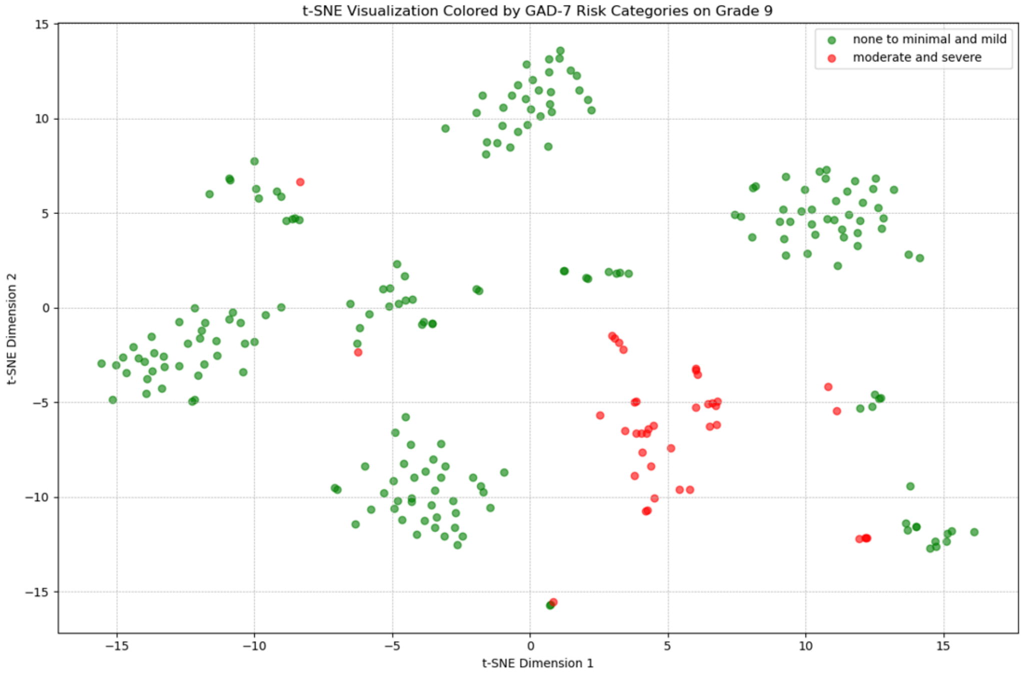

Figure 4.4: t-SNE representation of the risk group defined by GAD-7 within the grade 9 students ($n = 251$). This representation shows the cluster that differentiates, although not perfectly, between no-risk ($n = 212$, green clusters) and high-risk students ($n = 39$, red cluster in the middle right with outliers around the cluster mixing with green clusters).

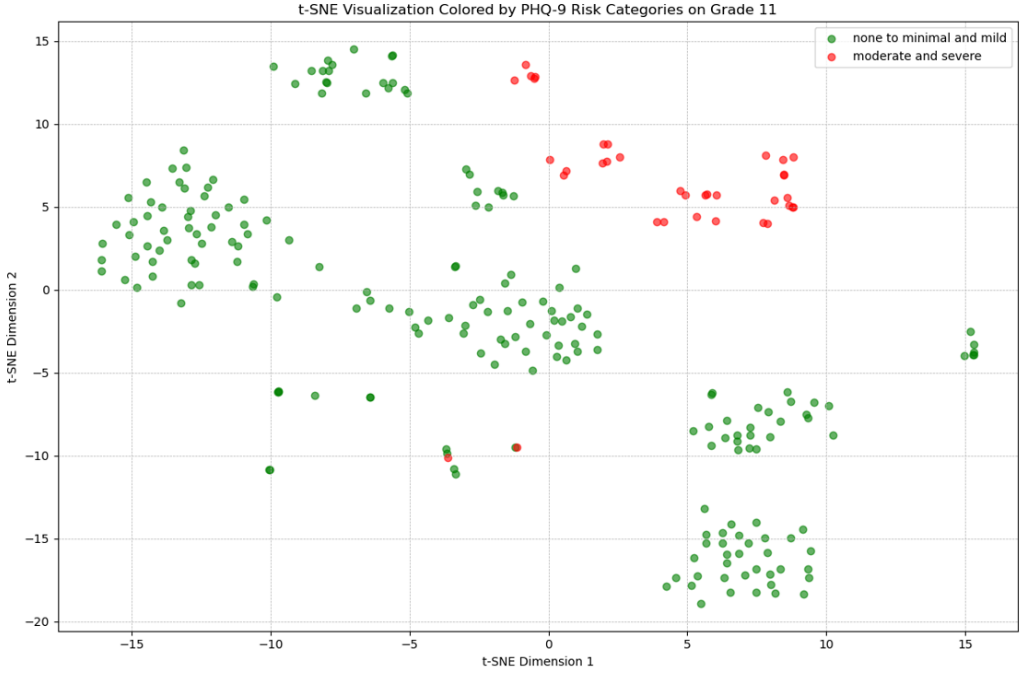

Figure 4.5: t-SNE representation of the risk group defined by PHQ-9 within the grade 11 students ($n = 242$). This representation shows the cluster that differentiates, although not perfectly, between no-risk ($n = 207$, green clusters) and high-risk students ($n = 35$, red cluster in the middle top with outliers mixing in the bottom small green cluster).

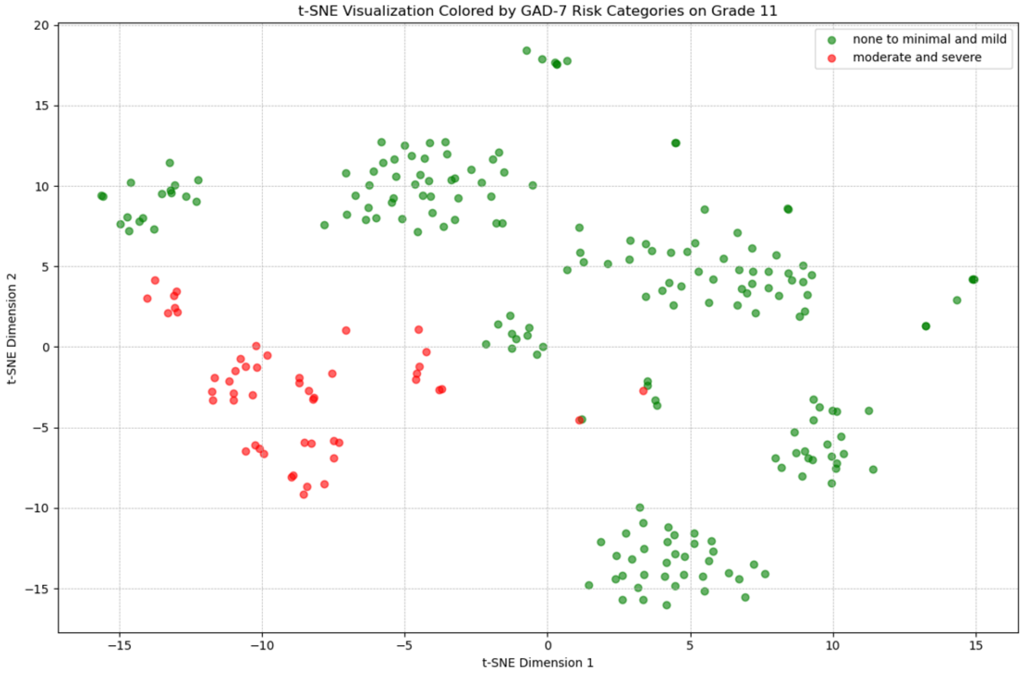

Figure 4.6: t-SNE representation of the risk group defined by GAD-7 within the grade 11 students ($n = 242$). This representation shows the cluster that differentiates, although not perfectly, between no-risk ($n = 192$, green clusters) and high-risk students ($n = 50$, red cluster in the middle left with outliers mixing in the middle small green cluster).

Based on these findings, the next step in the research will involve the development and validation of a machine learning model that can efficiently classify students into appropriate risk categories. The model's performance will be evaluated based on its accuracy, sensitivity, and specificity in predicting high-risk students.

4.3 Feature Importance of Random Forests

In this phase of the study, feature importance is determined by evaluating each feature's contribution to model accuracy within the random forest algorithm, which includes numerous decision trees. This assessment is stratified by gender and includes a focus on the available demographic categories within the dataset. Due to the limited sample sizes for demographic groups beyond White and Asian, these categories are not included in the current analysis. Future studies should aim to include a more diverse demographic representation to ensure comprehensive and inclusive modeling.

4.3.1 Middle School

Features indicating comorbidity, such as responses to questions about death and the T scores for anxiety or depression, stand out as the most influential for distinguishing between risk and no-risk groups among middle school students when using additional comorbidity features. The core class average emerges as a key academic feature that consistently affects the comorbidity indicators. Furthermore, when considering the variable of race, absences from school also appear to be a significant academic factor contributing to comorbidity, indicating a potential area of focus for early identification and intervention.

Table 4.1: Top three feature Contributions in Anxiety T Score Prediction with Risk Scores and Student Statuses in Middle School Students (Grade 7 and Grade 8)

		Question about death	Depression T score	Core class average T1/T2	SPED status	504 Plan status	ELL status	Tardy	Dismissal	Absence
All			2	3				1		
Gender	F	2	1	3						
	M	3	1	2						
Race	A	1	3							2
	W	3	1							2

Table 4.2: Top three feature Contributions in Depression T Score Prediction with Risk Scores and Student Statuses in Middle School Students (Grade 7 and Grade 8)

		Question about death	Anxiety T score	Core class average| T1/T2	SPED status	504 Plan status	ELL status	Tardy	Dismissal	Absence
All		3	1	2						
Gender	F	2	1	3						
	M	2	1	3						
Race	A		1	2						3
	W	2	1						3	

Table 4.3: Top three feature Contributions in Question About Death Prediction with Risk Scores and Student Statuses in Middle School Students (Grade 7 and Grade 8)

		Depression T score	Anxiety T score	Core class average| T1/T2	SPED status	504 Plan status	ELL status	Tardy	Dismissal	Absence
All		2	1	3						
Gender	F	2	1	3						
	M	3	1	2						
Race	A	2	1	3						
	W	1	2	3						

4.3.2 High School

In high school students, the trend observed in middle school is largely mirrored. However, it is notable that the response to question 9 is not the primary comorbidity indicator. Instead, the PHQ-9 and GAD-7 scores consistently exhibit strong predictive power. Academic performance, as indicated by average Q1/Q2 scores, remains a consistently influential factor. Furthermore, when the data is stratified by gender, absences become one of the most significant predictors. Similarly, when analyzed by race, tardiness emerges as a leading feature, suggesting that these school attendance-related factors are important in the context of comorbidities in high school students.

Table 4.4: Top three feature Contributions in GAD-7 Prediction with Risk Scores and Student Statuses in High School Students (Grade 9 and Grade 11)

		Endorse	PHQ-9	Average	SPED	504 Plan	ELL	Tardy	Dismissal	Absence
		question 9	total	Q1/Q2	status	status	status			
All			1	2						3
Gender	F		1	2						3
	M		1	2						3
Race	A		1	2				3		
	W		1	2				3		

Table 4.5: Top three feature Contributions in PHQ-9 Prediction with Risk Scores and Student Statuses in High School Students (Grade 9 and Grade 11)

		Endorse	GAD-7	Average	SPED	504 Plan	ELL	Tardy	Dismissal	Absence
		question 9	total	Q1/Q2	status	status	status			
All		2	1	3						
Gender	F		1	2						3
	M		1	2						3
Race	A		1	2				3		
	W		1	2				3		

Chapter 5

Conclusion and Discussion

5.1 Conclusion

This work uses academic and psychometric data to construct and refine random forest classifiers for the precise prediction of mental health risks among middle and high school students in Westborough. The *t*-SNE analysis offers promising indications that, with the current covariates, a model of reasonable performance is achievable. Furthermore, it has been observed that clusters become less distinct as student grade levels increase, implying that the complexity of predictions increases with age.

Through the implementation of a two-way ANOVA, the study examined the impact of gender and race/ethnicity on mental health outcomes. The results highlighted significant variances in depression and anxiety scores associated with these demographic variables, thus indicating their influence on mental health risks.

For middle schoolers, gender is a crucial predictor of both Depression T and Anxiety T scores. This denotes more pronounced symptoms in one gender compared to the other. Race/ethnicity was similarly influential, with notable differences in mental health outcomes across various groups. The analysis did not reveal any significant interaction effects between gender and race/ethnicity, suggesting that the effects of these factors are distinct and independent.

The trend continued in the high school cohort with even greater gender disparities observed in GAD-7 and PHQ-9 scores. Female students reported higher levels of anxiety and depression than males. Additionally, racial/ethnic discrepancies were apparent, especially in the domain of anxiety, with White students reporting higher levels than Asian students.

Looking at the feature importance, the consistent top predictors are psychometric tests that were not used for classifying the risk group, indicating the comorbidity and predictive power of the test for predicting another psychometric test. Moreover, it is consistent in both middle and more pronounced high schools that academic performance, like core class average and average scores on each quarter, are the main academic features that are one of the top features contributions. This indicates that stress on academic performance needs to be looked into. Not only that, but also it is also worth mentioning that absences are also a pronounced feature in high school students that influence the prediction of the model.

5.2 Discussion

The outcomes of this investigation reveal a nuanced relationship between gender, race/ethnicity, and the mental health of adolescents. The data highlights that these demographic variables independently exert significant effects on the manifestation of depression and anxiety symptoms across both middle and high school student groups, with no notable interaction effects. This is in concordance with existing studies, such as those by Afroz et al. (2023), which postulate that the impact of demographic factors on mental health may be attributed to the associated social, cultural, and biological influences. Additionally, the gradation by school year serves as a proxy for age, suggesting an increase in the complexity of mental health assessments in relation to environmental influences as students mature.

The analytical results derived from the random forest models consistently position academic performance as a primary covariate, with its influence being particularly pronounced among high school students. Furthermore, attendance patterns such as absenteeism emerge as significant indicators of mental health, corroborating findings from the existing literature (Finning et al., 2019; Lawrence et al., 2019). For instance, gender-conditioned absenteeism in high school and race-conditioned absenteeism in middle school are among the factors closely aligned with mental health issues.

Looking at the problem that has been analyzed, the following list are the suggested data-driven action for tackling mental health issue in Westborough middle and high school.

1. Personalized Interventions: Develop targeted interventions that are sensitive to the demographic variables of gender and race/ethnicity. This could include culturally adapted counseling services and gender-specific mental health programs. Although, as limited by the number of samples in this study, more research on the interventions and covariates' influences on minorities is needed.

2. Academic Support Systems: As academic performance is consistently a leading covariate, recognizing the link between academic performance and mental health, schools should implement support structures that address academic pressures, such as tutoring programs and stress management workshops.

3. Attendance Monitoring: Regular monitoring of student attendance and the implementation of early intervention strategies for students with high absenteeism or tardiness could serve as an early warning system for potential mental health concerns.

4. Age-Specific Strategies: As shown in t-SNE result, the distribution of the risk, and no-risk group is less distinctive in the higher age. Adjust mental health screening and intervention programs to align with the developmental stages of students, recognizing that the indicators and expressions of mental health issues may evolve with age.

5.3 Future Works

This work is a partial expansion of Lo (2023). Due to the limitation of the data, specifically on the score for each question in the questionnaire, the full analysis could not be done similarly. More detailed analyses are needed to conclude the work. Moreover, more robust evaluation metrics for the feature importance or statistical analysis on the feature importance, for example, permutation test, can be done to reinforce the conclusion further.

Bibliography

Afroz, N., Kabir, E., and Alam, K. (2023). A latent class analysis of the socio-demographic factors and associations with mental and behavioral disorders among australian children and adolescents. *PLoS One*, 18(5):e0285940.

Biau, G. (2012). Analysis of a random forests model. *Journal of Machine Learning Research*, 13:1063–1095.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.

Breiman, L. (2001). Random forest. *Machine Learning*, 45(1):5–32.

Castellanos-Ryan, N., Brière, F. N., O'Leary-Barrett, M., Banaschewski, T., Bokde, A., Bromberg, U., Büchel, C., Flor, H., Frouin, V., Gallinat, J., Garavan, H., Martinot, J.-L., Nees, F., Paus, T., Pausova, Z., Rietschel, M., Smolka, M. N., Robbins, T. W., Whelan, R., Schumann, G., Conrod, P., and IMAGEN Consortium (2016). The structure of psychopathology in adolescence and its common personality and cognitive correlates. *J. Abnorm. Psychol.*, 125(8):1039–1052.

Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., and Francis, S. E. (2000). Assessment of symptoms of dsm-iv anxiety and depression in children: a revised child anxiety and depression scale. *Behaviour Research and Therapy*, 38(8):835–855.

de Figueiredo, C. S., Sandre, P. C., Portugal, L. C. L., de Oliveira, T. M., da Silva Chagas, L., Ícaro Raony, Ferreira, E. S., de Araujo, E. G., dos Santos, A. A., and Bomfim, P. O.-S. (2021). Covid-19 pandemic impact on children and adolescents' mental health: Biological, environmental, and social factors. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 106:110171.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1 – 26.

Eun, K. Y. and Boram, L. (2019). The psychometric properties of the patient health questionnaire-9 in a sample of korean university students. *Psychiatry Investig*, 16(12):904–910.

Finning, K., Ukoumunne, O. C., Ford, T., Danielson-Waters, E., Shaw, L., Romero De Jager, I., Stentiford, L., and Moore, D. A. (2019). Review: The association between anxiety and poor attendance at school - a systematic review. *Child Adolesc Ment Health*, 24(3):205–216.

Gaylor, E. M., Krause, K. H., Welder, L. E., Cooper, A. C., Ashley, C., Mack, K. A., Crosby, A. E., Trinh, E., Ivey-Stephenson, A. Z., and Whittle, L. (2023). Suicidal thoughts and behaviors among high school students - youth risk behavior survey, united states, 2021. *MMWR Suppl*, 72(1):45–54.

Ingram, R. E. and Luxton, D. D. (2005). *Vulnerability-Stress Models.*, pages 32–46. Development of psychopathology: A vulnerability-stress perspective. Sage Publications, Inc, Thousand Oaks, CA, US.

James, G., Witten, D., Hastie, T., Tibshirani, R., and Taylor, J. (2023). *An Introduction to Statistical Learning: with Applications in Python*. Springer International Publishing.

Kroenke, K., Spitzer, R. L., and Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*, 16(9):606–613.

Kučak, D., Juričić, V., and Đambić, G. (2018). Machine learning in education-a survey of current research trends. *Annals of DAAAM & Proceedings*, 29.

Lawrence, D., Dawson, V., Houghton, S., Goodsell, B., and Sawyer, M. G. (2019). Impact of mental disorders on attendance at school. *Australian Journal of Education*, 63(1):5–21.

Lipson, S. K., Zhou, S., Abelson, S., Heinze, J., Jirsa, M., Morigney, J., Patterson, A., Singh, M., and Eisenberg, D. (2022). Trends in college student mental health and help-seeking by race/ethnicity: Findings from the national healthy minds study, 2013–2021. *Journal of Affective Disorders*, 306:138–147.

Lo, W. (2023). Westborough high school mental health predictive analyses. Technical report, Worcester Polytechnic Institute.

Maddux, J. E. and Winstead, B. A. (2020). *Psychopathology: Foundations for a Contemporary Understanding*. Routledge, Milton, 5th edition.

Marques de Miranda, D., da Silva Athanasio, B., Sena Oliveira, A. C., and e Silva, A. C. S. (2020). How is covid-19 pandemic impacting mental health of children and adolescents? *International Journal of Disaster Risk Reduction*, 51:101845.

Pongpaichet, S., Nirunwiroj, K., and Tuarob, S. (2022). Automatic assessment and identification of leadership in college students. *IEEE Access*, 10:79041–79060.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., and Löwe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine*, 166(10):1092–1097.

Tate, A. E., McCabe, R. C., Larsson, H., Lundström, S., Lichtenstein, P., and Kuja-Halkola, R. (2020). Predicting mental health problems in adolescence using machine learning techniques. *PLOS ONE*, 15(4):1–13.

Thieme, A., Belgrave, D., and Doherty, G. (2020). Machine learning in mental health: A systematic review of the hci literature to support the development of effective and implementable ml systems. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 27(5):1–53.

Usher, W. and Curran, C. (2017). Predicting Australia's university students' mental health status. *Health Promotion International*, 34(2):312–322.

Appendices

Appendix A

Model Performance

Table A.1: 10-fold Cross-validation Random Forest Performance on the Non-stratified Samples

School	Target	Precision	Recall	F1
Middle School (n = 519)	Depression T Score	0.8153	0.8245	0.8151
	Anxiety T Score	0.7018	0.7306	0.7060
	Question about Death	0.7069	0.6721	0.6789
High School (n = 493)	PHQ-9	0.7664	0.6535	0.6815
	GAD-7	0.7424	0.6143	0.6359
	Question 9	0.6528	0.6222	0.6285

Table A.2: 10-fold Cross-validation Random Forest Performance on the Gender Stratified Samples

School	Target	Precision	Recall	F1
Male, Middle School (n = 271)	Depression T Score	0.7646	0.7460	0.7450
	Anxiety T Score	0.7407	0.7462	0.7433
	Question about Death	0.7366	0.6784	0.6903
Female, Middle School (n = 248)	Depression T Score	0.8077	0.7558	0.7547
	Anxiety T Score	0.7230	0.6964	0.6823
	Question about Death	0.6855	0.6444	0.6473
Male, High School (n = 222)	PHQ-9	0.4589	0.4950	0.4761
	GAD-7	0.7681	0.7121	0.7137
	Question 9	0.5181	0.5476	0.5323
Female, High School (n = 271)	PHQ-9	0.8178	0.7273	0.7498
	GAD-7	0.6170	0.5814	0.5896
	Question 9	0.8557	0.6895	0.7330

Table A.3: 10-fold Cross-validation Random Forest Performance on the Race Stratified Samples

School	Target	Precision	Recall	F1
White, Middle School (n = 308)	Depression T Score	0.8599	0.8778	0.8601
	Anxiety T Score	0.7281	0.7595	0.7319
	Question about Death	0.7139	0.6631	0.6734
Asian, Middle School (n = 185)	Depression T Score	0.5267	0.5410	0.5336
	Anxiety T Score	0.8417	0.8500	0.8457
	Question about Death	0.7220	0.6667	0.6687
White, High School (n = 283)	PHQ-9	0.7763	0.6836	0.7108
	GAD-7	0.7377	0.7051	0.7034
	Question 9	0.8060	0.7108	0.7380
Asian, High School (n = 166)	PHQ-9	0.6103	0.5967	0.5943
	GAD-7	0.4610	0.4942	0.4763
	Question 9	0.4603	0.4933	0.4762