Westborough Public Schools Mental Health Data Analysis

A Major Qualifying Project
Submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirements for the
Degree of Bachelor of Science
in Computer Science

By

Ben Kresge
Aukkawut Ammartayakun
Khoi Pham
Brian Borghesani

Advisors
Benjamin Nephew
Dmitry Korkin

Co-Advisor
Angela Incollingo Rodriguez

Abstract

In response to increasing concern over adolescent well-being in recent years, Westborough Public Schools (WPS) began collecting mental health data through self-report surveys in 2022-2023, in an effort to identify at-risk students. Performing such collection can be difficult and expensive, so accurate predictions using existing data would be potentially useful for WPS and other schools. This provides an opportunity for the application of Computer Science expertise to meet an immediate and practical social need. While this work is a continuation of a previous project, there were some significant differences: the data include a wider demographic range (students at additional ages/grade levels), but lack question specific resolution for mental health. Using the machine learning technique of Random Forest Classifiers (RFCs) and statistical analysis including ANOVA, some important features and trends were identified that align with existing research.

Background

In their continued commitment to supporting mental health in students, Westborough Public Schools collected mental health, demographic, and academic data. In 2022, only data from the 9th grade was collected, which was expanded to two high school and two middle school grades for 2023. The original collection was prompted in part by the 2021 MetroWest Adolescent Health Survey which found stress, anxiety, depression, and associated risk behaviors such as substance use and violence to be increasing in the adolescent student population in Massachusetts. Further cause for concern came from the Covid-19 pandemic which has had a negative impact on mental health in adolescents (de Figueiredo et al., 2021; Marques de Miranda et al., 2020, as cited in Ammartayakun et al., 2024).

Several cohorts of WPI students are working with WPS mental health data under the auspices of Professors Nephew and Korkin. Last year's work (Lo, 2023) showed the effectiveness and utility of RFC models in predicting mental health risk in WPS students. While the performance was not exceptional, the models could potentially aid WPS in the early identification of at-risk students. Additionally, Lo (2023) was able to specify a subset of important features to be considered. This year's work serves to further verify the utility of such models and extends the scope to middle school students.

Mental Health Data

The main targets for mental health measurement were anxiety and depression risk, along with suicidal ideation. Mental health data were collected using a different set of surveys for middle school (7th and 8th grade) and high school (9th and 11th grade) students due to different surveys being more appropriate for different age groups. All the surveys administered cover the relevant current criteria in the DSM-5, an American Psychiatric Association publication that is recognized as the standard for diagnostic criteria across mental health disciplines. Important to note however, is that the surveys are all face-valid, meaning it is clear to the student exactly what is being tested for and how their responses will roughly contribute to the final result. The consequence being that if a student wanted to appear more or less at risk, it is possible to answer in such a way as to achieve that result.

Middle School

The shortened 25 question form of the Revised Children's Anxiety and Depression Scale (RCADS) was administered to middle school students. The full set contains 47 questions and has subscales including obsessive compulsive disorder and separation anxiety whereas the 25 question set only has the anxiety score, depression score, and combined score (Ebesutani et al., 2012). Question 18 ("I think about death") was used as a comparison point for suicidal ideation, even though the wording of the question suggests a more general conceptualization of death rather than suicidal ideation per se.

High School

The depression risk target was measured using the Patient Health Questionnaire (PHQ-9) and suicidal ideation was measured using Question 9 ("Thoughts that you would be better off dead, or of hurting yourself"), while the anxiety target was measured using the Generalized Anxiety Disorder scale (GAD-7). Both of these surveys are intended for adult populations, but besides slight differences in optimal score cutoffs, they still retain much of their useful psychometric properties when applied to adolescents (Richardson et al., 2010; Mossman et al., 2017).

Methodology

The data were split between middle and high school since that corresponds to the means of data collection. It was found that there was no meaningful difference between grades within these groups and combining the grades allowed for a larger training set and sample size. Unless otherwise stated, middle school refers to the combination of 7th and 8th grade, and high school refers to the combination of 9th and 11th grade. The schools were further split into subgroups by race/ethnicity and gender. It was not possible to perform analysis on the intersections of these subgroups (e.g. white males) due to too few positive results for the risk targets. Black and non-binary students had to be excluded as subgroups due to sample size constraints, although they were not removed from the dataset and were factored into all groups to which they should otherwise belong. Similarly, data from students who reported multiple races/ethnicities were not included in the race/ethnicity subgroups but were included in the overall population and gender groups.

The models used for prediction were Random Forest Classifiers (an ensemble method using decision trees as estimators). The parameters for the RFCs were as follows: n_estimators=10, max_depth=21, min_samples_split=2, min_samples_leaf=1. To calculate the performance of the models, k-fold validation was used (with k=10). Instead of predicting the score value, the model attempted to categorize targets into risk binaries. The binary was divided by literature accepted cutoffs of 9 for PHQ-9 and GAD-7, 65 for RCADS-25 anxiety and depression T-scores, and any non-zero response for PHQ-9 Q9 and RCADS-25 Q18. This means that a positive categorization corresponds to a level of clinical significance. Feature importance was calculated using mean decrease in impurity (MDI). These rankings were checked against the drop in F1-score and accuracies when the top features were removed.

The models and analyses were run using Python scripts. The data were read, formatted, and structured using pandas and numpy. The models used were RandomForestClassifiers from sklearn, which also provided utilities such as k-fold cross-validation. ANOVA was run via pingouin, and plots were created using matplotlib or Excel.

Results

ANOVA Findings

Two-way ANOVAs were run between gender and race/ethnicity and outcome risk scores, with main (ME) and interaction effects (IE) provided.

Middle School

Depression T-score showed a significant difference across gender (ME) ($F=20.21$, $p=8.64e-06$). The scores were higher for females compared to males. There was also a significant effect of race/ethnicity (ME) ($F=7.05$, $p=9.61e-04$), with White students having higher depression scores. The interaction between gender and race/ethnicity did not significantly influence depression T-scores ($F=0.31$, $p=0.731$). Anxiety T-scores exhibited significant differences across gender ($F=35.74$, $p=4.30e-09$), again being higher in females. There was a significant effect of race/ethnicity on anxiety T-scores ($F=4.82$, $p=8.41e-03$), with White students again having higher scores. The interaction between gender and race/ethnicity did not significantly influence anxiety T-scores ($F=2.02$, $p=0.133$). Question 18 did not show a significant difference across gender ($F=2.42$, $p=0.12$) or race/ethnicity ($F=0.29$, $p=0.745$). The interaction between gender and race/ethnicity also did not significantly influence Question 18 ($F=0.09$, $p=0.915$).

High School

GAD-7 scores exhibited significant differences across race/ethnicity ($F=4.9$, $p=0.03$) and even more so across gender ($F=26.0$, $p<0.01$). Anxiety levels were higher in White students compared to Asian students and higher in females than in males. There was no significant interaction between gender and race/ethnicity ($F=0.1$, $p=0.7$). As for PHQ-9, scores differed significantly across gender ($F=17.7$, $p<0.001$), with depression being more common in females than in males. However, there was no effect of race/ethnicity ($F=1.1$, $p=0.3$) and no significant interaction ($F=0.6$, $p=0.44$). Question 9 was not significantly different across gender ($F=1.6$, $p=0.21$), race/ethnicity ($F=0.0$, $p=0.94$), or gender and race/ethnicity combined ($F=0.0$, $p=0.98$).

Model Performances

The performance of the RFCs are provided in Figures 1-8, and in table form in the Appendix (Tables 9-16). The accuracy for most targets in both schools were usually in the 0.85-0.95 range, with the exception of 0.73 for the Q18 target in middle school. However, accuracy is most likely an overestimation of performance due to the unbalanced nature of the data. A hypothetical model could achieve reasonably good accuracies by always categorizing each student into the low risk category. The F1 score helps to account for this issue, but a balanced accuracy is also provided which accounts for the imbalance in the data. The balanced accuracies in models across the entire school are mostly in the 0.6-0.7 range. Some variation between subgroups can be explained by the limited positive case data (Appendix Tables 7,8). Overall, the models performed slightly better for middle school students than high school students.

Middle School

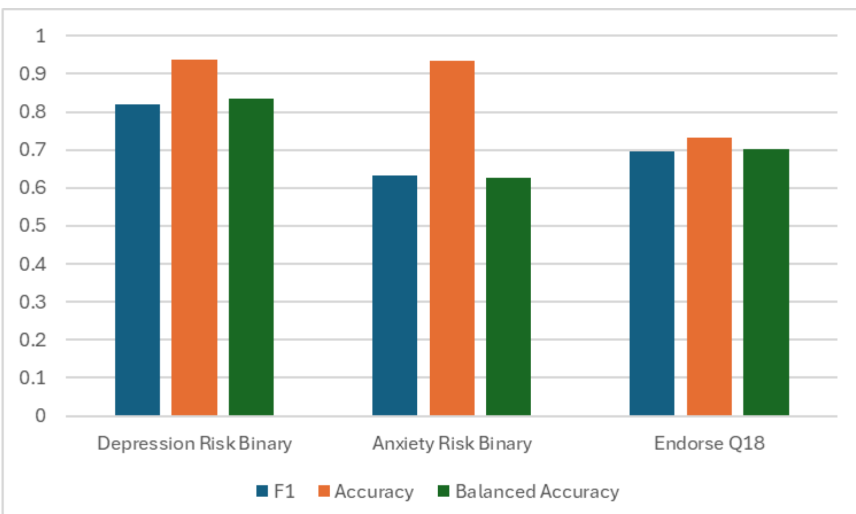

Figure 1 -RFC Performances in All Middle School Students

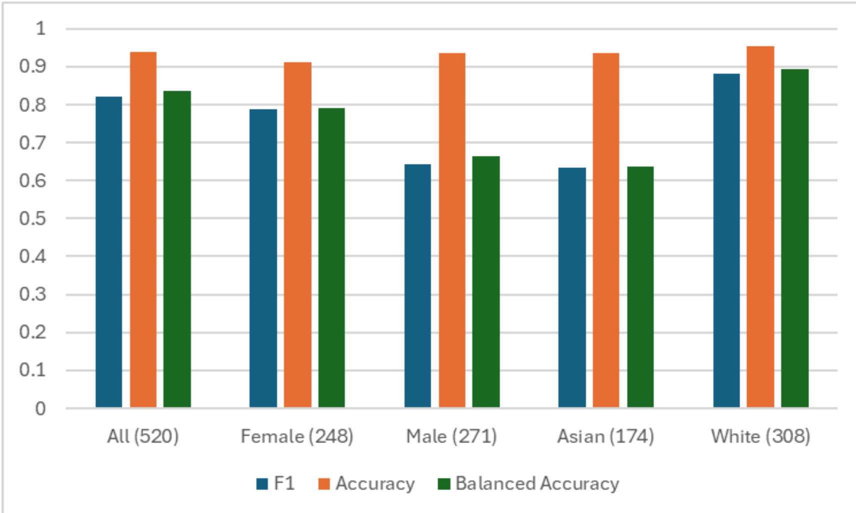

Figure 2 - RFC Performance for Depression Risk Prediction Across Gender and Race/Ethnicity Subgroups of Middle School Students (N Students in Subgroup)

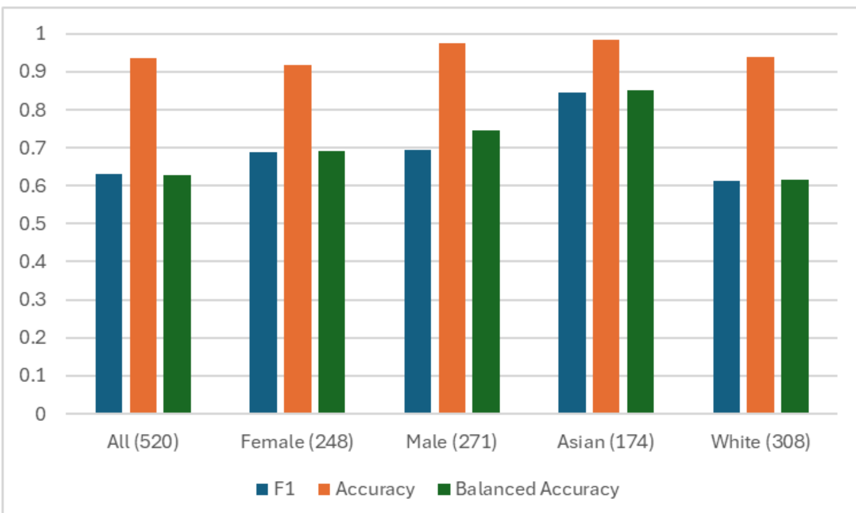

Figure 3 - RFC Performance for Anxiety Risk Prediction Across Gender and Race/Ethnicity Subgroups of Middle School Students (N Students in Subgroup)

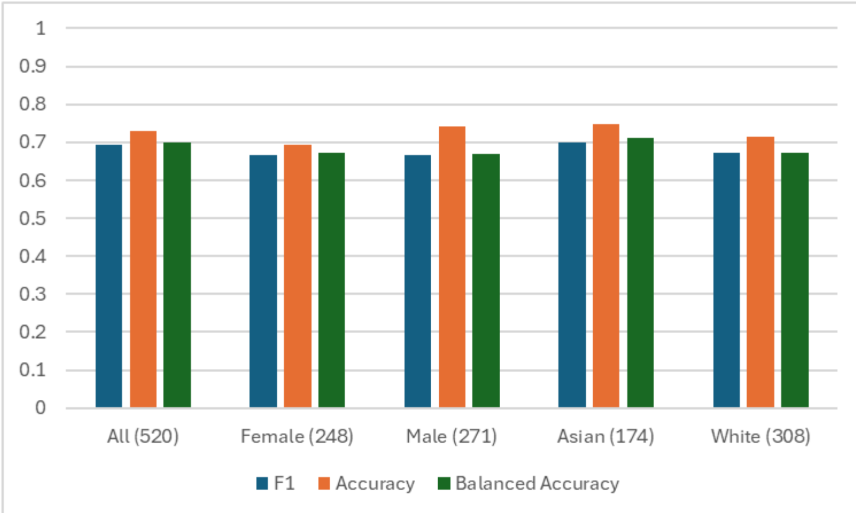

Figure 4 - RFC Performance for Endorse Q18 ("Think about death") Prediction Across Gender and Race/Ethnicity Subgroups of Middle School Students (N Students in Subgroup)

High School

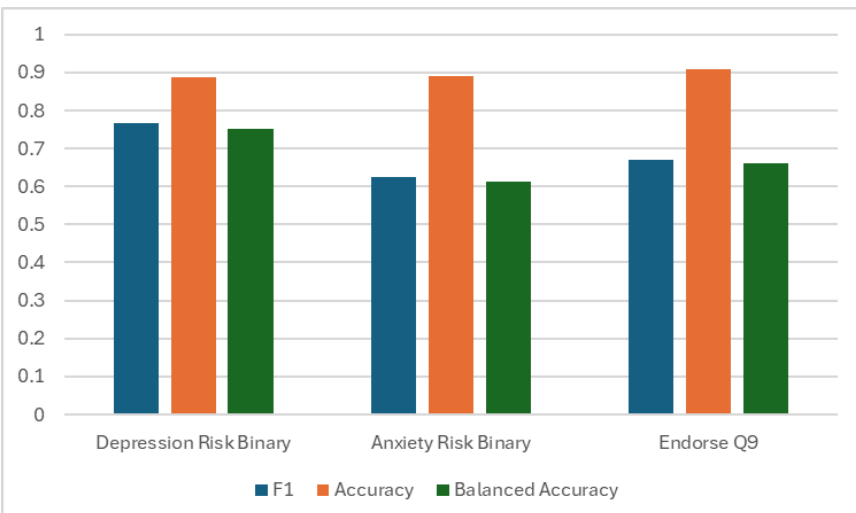

Figure 5 - RFC Performances in All High School Students

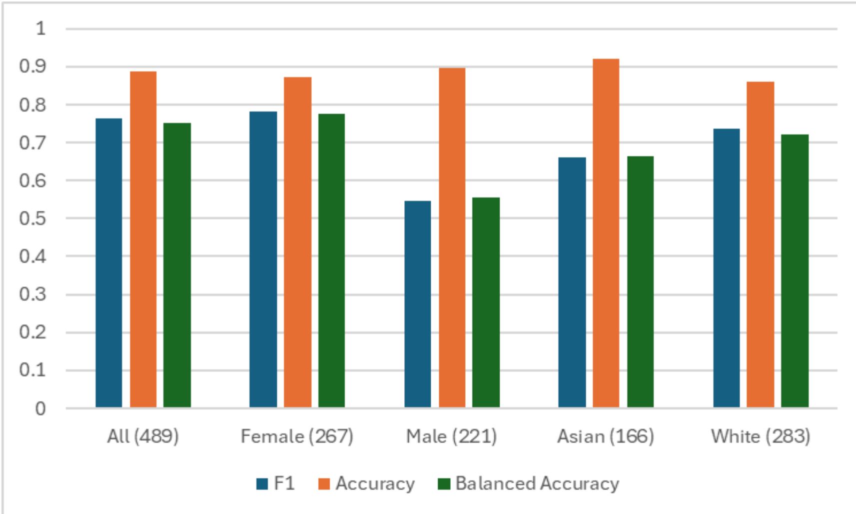

Figure 6 - RFC Performance for Depression Risk Prediction Across Gender and Race/Ethnicity Subgroups of High School Students (N Students in Subgroup)

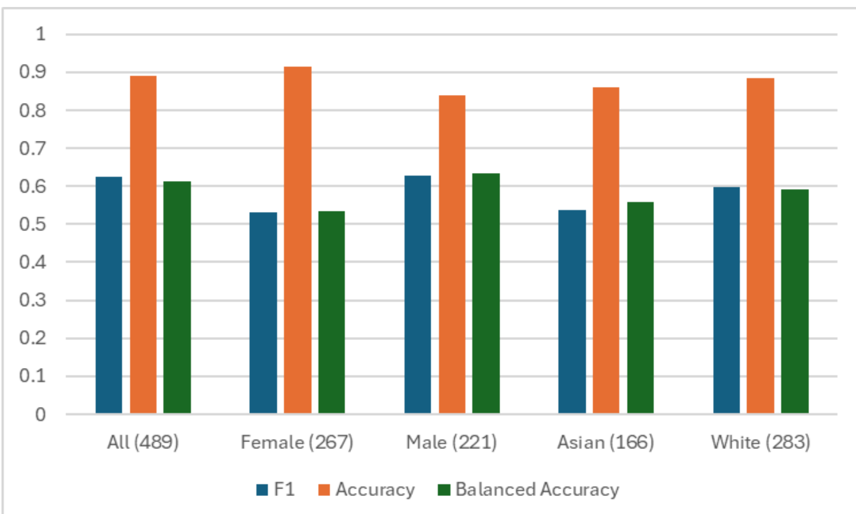

Figure 7 - RFC Performance for Anxiety Risk Prediction Across Gender and Race/Ethnicity Subgroups of High School Students (N Students in Subgroup)

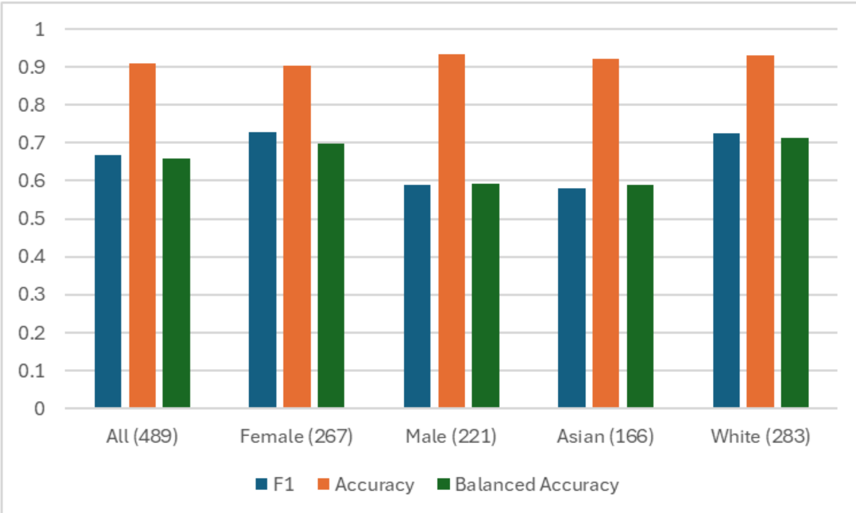

Figure 8 - RFC Performance for Endorse Q9 (“Thoughts that you would be better off dead”) Prediction Across Gender and Race/Ethnicity Subgroups of High School Students (N Students in Subgroup)

Feature Importance

The seven most important features by mean decrease in impurity (MDI) for each target are provided in Figures 9-14. After analyzing changes in model performance after feature removal, it appears to be difficult to interpret the significance of features with a MDI under approximately 0.15. While some features with a MDI < 0.15 lower performance metrics to a significant degree when removed, most lower the metrics only slightly, while some even increase metrics. For this reason, this paper will refer to features with a MDI of > 0.15 as "significant" and MDI ≤ 0.15 as "insignificant". Since this cutoff is somewhat arbitrary, features with a MDI less than 0.15 are still included in the figures and tables.

Middle School

Depression

The top predictor for depression across gender and race/ethnicity was always the anxiety T-score. Endorsing Q18 was an additional important feature for the White and Male subgroups, whereas grade average was an important feature for the Asian subgroup.

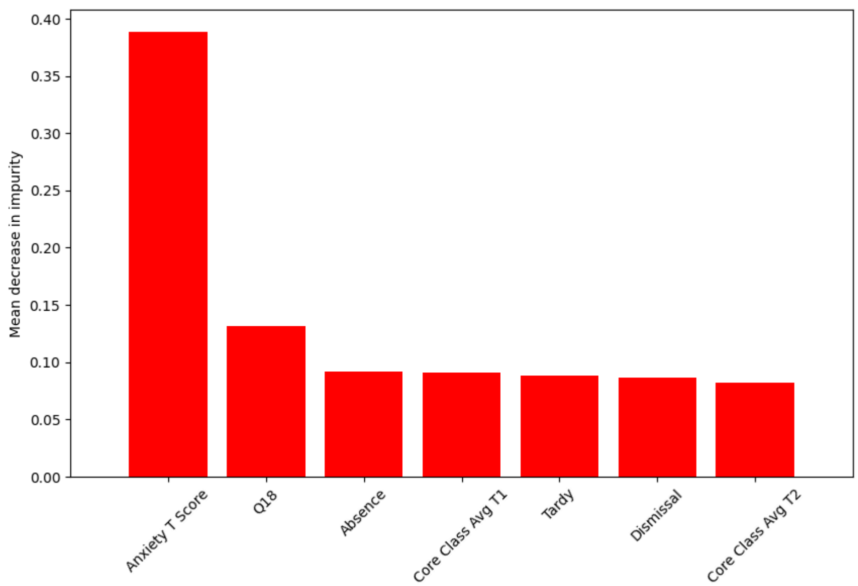

Figure 9 - Feature Importances for Depression Risk Prediction in All Middle School Students

		Anxiety T-Score	Endorse Q18	Grade Average (T1/T2)*	Absences	Tardies
Gender	F	0.354 (1)	0.146	0.101	0.080	0.061
	M	0.436 (1)	0.185 (2)	0.079	0.079	0.041
Race	A	0.288 (1)	0.123	0.203 (2)	0.069	0.091
	W	0.357 (1)	0.208 (2)	0.089	0.077	0.060

** The higher MDI feature between Core Class Avg 1 and 2*

Table 1 - MDI of Top Features by Subgroup for Depression Risk Prediction in Middle School Students

Anxiety

The top predictor for anxiety was the depression T-score for all subgroups besides Asian students, where it was barely surpassed by endorsing Q18. Additionally, dismissals were a significant feature for Asian students. Endorsing Q18 was a significant feature for the White subgroup, but barely missed the significance threshold for the Male and Female subgroups. Grade average was usually in the top three or four important features, but did not surpass the significance threshold in any subgroups and performance metrics were not reduced with its removal.

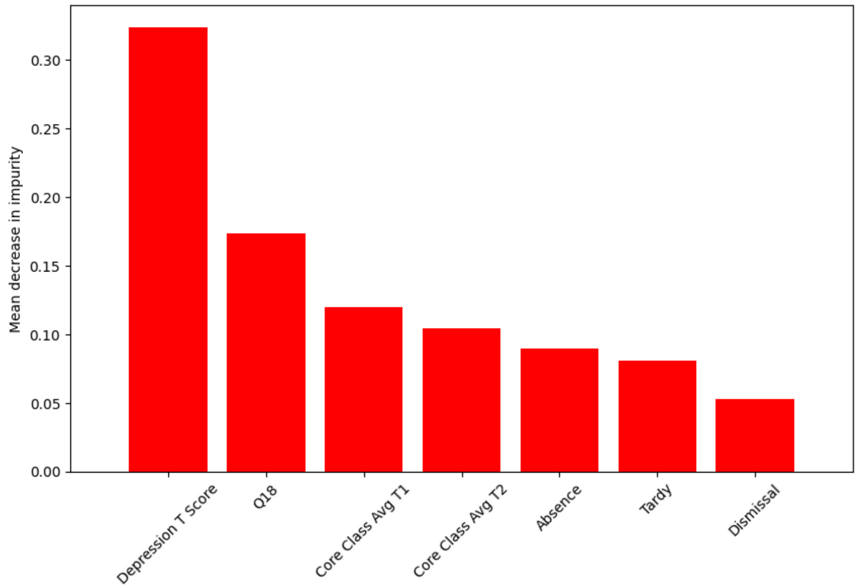

Figure 10 - Feature Importances for Anxiety Risk Prediction in All Middle School Students

		Depression T-Score	Endorse Q18	Dismissals	Grade Average (T1/T2)*	Absences
Gender	F	0.338 (1)	0.148	0.102	0.129	0.122
	M	0.230 (1)	0.149	0.120	0.134	0.075
Race	A	0.254 (2)	0.291 (1)	0.202 (3)	0.121	0.080
	W	0.310 (1)	0.226 (2)	0.093	0.094	0.077

** The higher MDI feature between Core Class Avg 1 and 2*

Table 2 - MDI of Top Features by Subgroup for Anxiety Risk Prediction in Middle School Students

Endorse Q18 ("Think about death")

The T-scores for anxiety and depression were both very important features for predicting Q18 endorsement. For the Female and Male subgroups the anxiety score was slightly more important, whereas the depression score was slightly more important for the Asian and White subgroups. Of the other features, while grade average almost met the threshold for significance for the White subgroup, for all other features and subgroups the importance was uncertain.

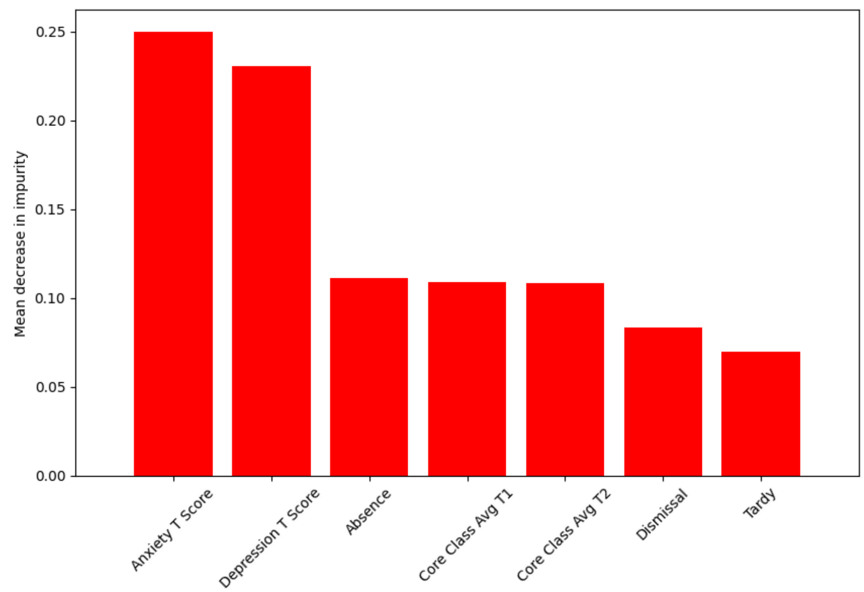

Figure 11 - Feature Importances for Endorse Q18 ("Think about death") Prediction in All Middle School Students

		Anxiety T-Score	Depression T-Score	Absences	Grade Average (T1/T2)*	Dismissals
Gender	F	0.270 (1)	0.226 (2)	0.101	0.107	0.094
	M	0.234 (1)	0.228 (2)	0.126	0.130	0.072
Race	A	0.283 (2)	0.315 (1)	0.111	0.082	0.055
	W	0.221 (2)	0.231 (1)	0.106	0.148	0.079

** The higher MDI feature between Core Class Avg 1 and 2*

Table 3 - MDI of Top Features by Subgroup for Endorse Q18 ("Think about death") Prediction in Middle School Students

High School

Westborough High School uses a course level system which is similar to the typical honors distinction. The feature "average course level" represents the average of the numeric course levels. The rough scale of the levels is 1-base course, 2/3-faster paced courses with more independence (college prep), 4-honors, and 5-either AP courses or equivalent acceleration if an AP course is not available.

Depression

The top and only significant predictor across all subgroups was the GAD-7 total score. Of the insignificant features, endorsing Q9 had a reasonably high MDI across all subgroups except Males. Grade average appears to have middling importance to all subgroups except White students, where the importance is lesser.

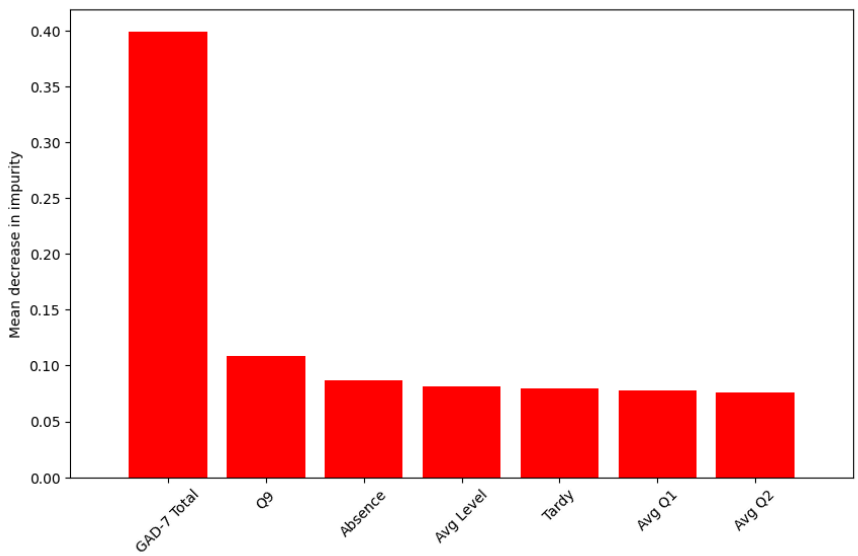

Figure 12 - Feature Importances for Depression Risk Prediction in All High School Students

		GAD-7 Total	Endorse Q9	Grade Average (Q1/Q2)*	Absences	Tardies	Average Course Level
Gender	F	0.335 (1)	0.114	0.104	0.094	0.092	0.084
	M	0.233 (1)	0.078	0.140	0.098	0.114	0.067
Race	A	0.361 (1)	0.141	0.130	0.074	0.109	0.077
	W	0.351 (1)	0.107	0.089	0.098	0.069	0.090

** The higher MDI feature between Grade Average for Q1 and Q2*

Table 4 - MDI of Top Features by Subgroup for Depression Risk Prediction in High School Students

Anxiety

The top predictor for anxiety risk was the PHQ-9 total for all subgroups. The grade average for Q2 specifically was a significant and the second most important feature for all subgroups except White students, where the average course level barely missed the significance threshold. In Asian students however, average course level was found to be significant. The grade average for Q1 was found to be a less important feature than in Q2 in all subgroups, which was only true for this target.

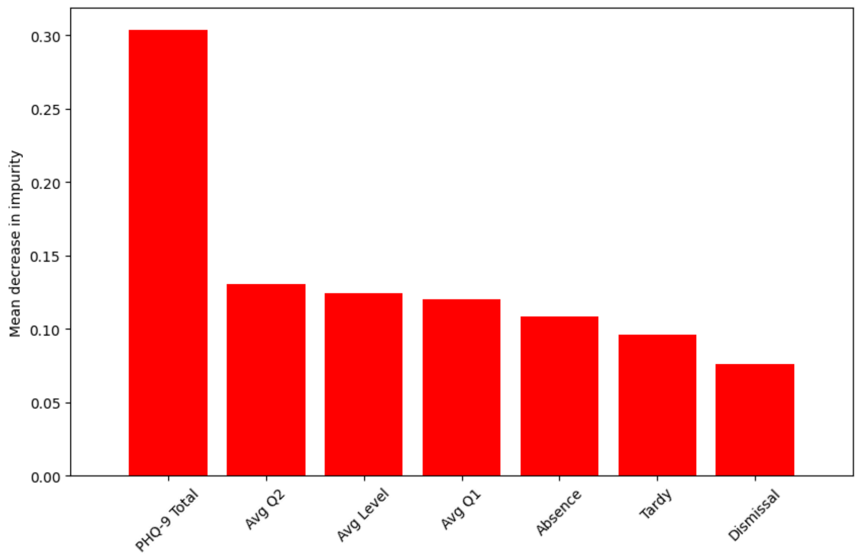

Figure 13 - Feature Importances for Anxiety Risk Prediction in All High School Students

		PHQ-9 Total	Grade Average Q2	Average Course Level	Absences	Grade Average Q1
Gender	F	0.287 (1)	0.163 (2)	0.106	0.112	0.108
	M	0.286 (1)	0.167 (2)	0.133	0.123	0.125
Race	A	0.224 (1)	0.210 (2)	0.184 (3)	0.118	0.128
	W	0.333 (1)	0.098	0.145	0.118	0.080

Table 5 - MDI of Top Features by Subgroup for Anxiety Risk Prediction in High School Students

Endorse Q9 ("Thoughts that you would be better off dead")

The top predictor for all subgroups was the PHQ-9 total. The GAD-7 total was the second most important feature for all subgroups except for Males, and was a significant feature for the Female and Asian subgroups. The feature importance of GAD-7 total for Males was surprisingly low, with more academic features such as average course level and grade average playing a more important role for that group.

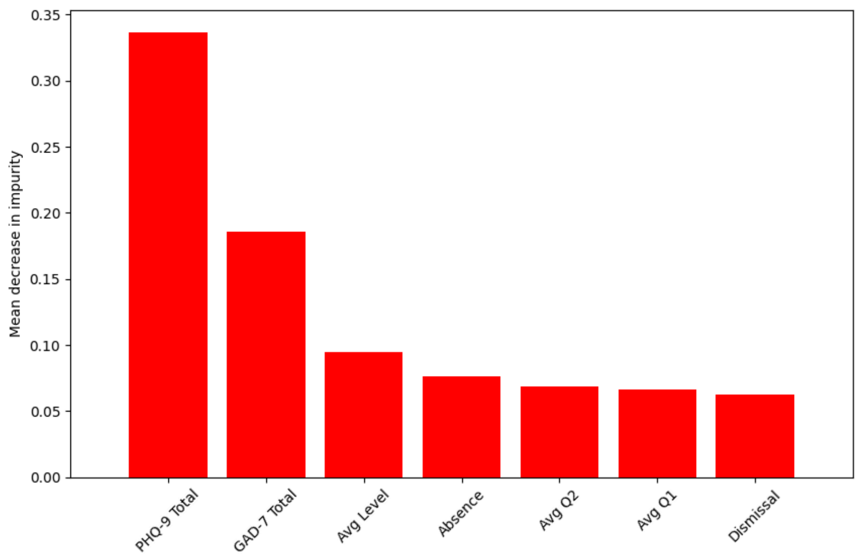

Figure 14 - Feature Importances for Endorse Q9 ("Thoughts that you would be better off dead") Prediction in All High School Students

		PHQ-9 Total	GAD-7 Total	Average Course Level	Absences	Grade Average (Q1/Q2)*
Gender	F	0.336 (1)	0.198 (2)	0.090	0.067	0.101
	M	0.272 (1)	0.063	0.146	0.073	0.130
Race	A	0.362 (1)	0.198 (2)	0.137	0.045	0.066
	W	0.332 (1)	0.117	0.077	0.108	0.097

* The higher MDI feature between Grade Average for Q1 and Q2

Table 6 - MDI of Top Features by Subgroup for Endorse Q9 ("Thoughts that you would be better off dead") Prediction in High School Students

Discussion

Mental health is influenced by a vast and complex array of interacting factors, which is only compounded by the fast pace and volatility of adolescence. The ANOVA tests demonstrate definite differences in prevalence and expression across subgroups of students, and the feature importances show variation in related factors. Lo (2023) also highlighted the importance of considering intersectionality when predicting mental health risk in the WPS population.

However, there are likely further factors and interactions that cannot be accounted for using the limited feature set. Segregating into groups may expose more truly random variance, or variance that is potentially influenced by factors that are not directly represented in the data, including, but not limited to the trajectory of school engagement (Li & Lerner, 2011), various community factors (Stirling et al., 2015), and school connectedness (Millings et al., 2012). There is also evidence for the existence of a "neglected group" of adolescents that do not fit into the traditional pattern at all (Antaramian et al., 2010).

That is not to say that there are not useful findings from this analysis. In agreement with Lo (2023) and Ammartayakun et al. (2024), the average grade of the student (in some form) was often a significant feature across all targets and groups. Even when not significant, it ranked reasonably high in MDI importance. Since grades are recorded and available at nearly all schools, they can be a useful tool for identifying possible at-risk students. After comparing the average grades across risk groups, it was found that lower average grades were associated with a higher mental health risk across all groups for all targets. The general association between poor grades and poor mental health is supported by literature (Burnett-Zeigler et al., 2012; Fröjd et al., 2007; Pascoe et al., 2019), however, in a systematic review, de Lijster et al. (2018) found that while anxiety disorders were correlated with increased feelings of academic impairment, the actual difference in average school results were mixed, further indicating the complexity of the system.

Another feature to consider when identifying at risk students is the absence count. Often considered a behavioral issue and investigated most at its extreme of truancy, non-attendance has a role to play in the interaction leading to, and/or in the consequences of, poor mental health (Lawrence et al., 2019). Again, this feature is of special interest since most school systems record these data. The general correlation from the data suggests that an increase in absence count is related to a decrease in mental health; the mean absence count for risk-positive students being at least 1-2 higher than risk-negative students across all schools, targets, and subgroups. While absence data were not available to Lo (2023), Ammartayakun et al. (2024) found absence count importance to be even higher than the results of this analysis suggest.

Finally, the most consistently important features were the mental health scales for the other targets; anxiety risk was best predicted by depression score, and vice versa. The models could be capturing a lurking factor common to both conditions, such as rumination (McLaughlin & Nolen-Hoeksema, 2011). There is evidence for anxiety being comorbid with depression and suicidal ideation (Maddux and Winstead, 2020, as cited in Ammartayakun et al., 2024). There is, however, another relationship the surveys share—they indicate willingness to self-report distress. If a student is willing to report distress in one survey, it is likely that they are willing to report distress in another. The correlation does not always hold; some students endorse questions on one survey but not the other, so it is important to take other factors into account, of which there are many (Colognori et al., 2012), such as grades and absences.

Limitations and Future Work

Feature Importance

Feature importance based on mean decrease in impurity may not be the best method for determining important factors contributing to mental health risk. MDI can be misleading for high cardinality features; it tends to have a bias towards features with many different values (Nguyen et al., 2015). Grade averages and middle school RCADS T-scores would be most vulnerable to this effect. Features such as ELL and 504 status would then be inaccurately represented as less important. A comparison to other methods of feature importance, such as permutation importance, may help distinguish true importance trends from this bias and provide more evidence of importance for features like average grades and the relationship between scales.

In one iteration of the model, one of the two grade average features was removed arbitrarily. The model performances were minimally altered, but the feature importance for the remaining grade average was significantly increased. It appears that the two features have some level of redundancy that is reflected in the feature importances. The true feature importances of the grade average may be higher than represented in the results of this analysis. Combining the features as a sum or average, or even using the difference over time may be a useful new constructed feature to examine in future work.

Diversity and Sample Size

While the dataset was large enough to run analyses, the sample sizes were small enough to be subject to high variance and certain subgroups were further limited by very small positive counts or rates (Appendix Tables 7,8). The consequence is that the results have limited applicability or reliability for some minorities in the population. This is also why the interaction between gender and race/ethnicity was unable to be explored using the RFC models. It was common to have fewer than 10 positive cases in these more specific groups which made the 10-fold cross-validation invalid.

Mental Health Data

Question level data were not available for this dataset. Lo (2023) found that the best performing models used the granular question level data. It stands to reason that a predictive model would perform better with more data, and for the specific use case, the question level data could allow inferences to be made about factors that are not measured, such as certain aspects of home life, or sleeping habits. Additionally, a higher cutoff for positive risk should be considered. Literature suggests that a threshold of around 11 for major depressive disorder on the PHQ-9 "was optimal for maximizing sensitivity without loss of specificity" (Richardson et al., 2010), and that the same cutoff may be appropriate for the GAD-7 (Mossman et al., 2017).

In conjunction with a change in threshold, it would also be informative to compare self-report data to diagnostic data. While the PHQ-9 and GAD-7 are well established screening measures, they do not necessarily reflect the actual prevalence of depression and anxiety in the population they are used for. Since the surveys are face-valid, the student is faced with the decision of whether or not to self-disclose. It may be the case that this population of students tends to over or under report their own symptoms.

Concluding Recommendation

From the results of this work and the others on WPS mental health data, it is evident that there is promise in using RFCs for predictive analysis. The next step could be to develop a deliverable that allows professionals in the WPS system to use RFC predictions to augment the process of identifying at-risk students. While the models can no doubt be further refined and enhanced with more features, there will be diminishing returns from the amount of effort needed to collect and process the data—there will always be variance and trends not captured in the data, and thus students that will be missed. Testing the model performance *in situ* would also be the most effective method towards improving the utility of future models. This is not to say that predictive models should in any way replace or take precedence over trained professionals, instead they could be used as a tool to help identify students for professional attention.

References

Ammartayakun, A., Kresge, B., Pham, K., & Mraz, L. (2024). *Westborough High School mental health predictive analysis* [Interactive qualifying project, Worcester Polytechnic Institute]. Digital WPI

Antaramian, S. P., Huebner, E. S., Hills, K. J., & Valois, R. F. (2010). A dual-factor model of Mental Health: Toward a more comprehensive understanding of youth functioning. *American Journal of Orthopsychiatry, 80*(4), 462–472. https://doi.org/10.1111/j.1939-0025.2010.01049.x

Burnett-Zeigler, I., Walton, M. A., Ilgen, M., Barry, K. L., Chermack, S. T., Zucker, R. A., Zimmerman, M. A., Booth, B. M., & Blow, F. C. (2012). Prevalence and correlates of mental health problems and treatment among adolescents seen in primary care. *Journal of Adolescent Health, 50*(6), 559–564. https://doi.org/10.1016/j.jadohealth.2011.10.005

Colognori, D., Esseling, P., Stewart, C., Reiss, P., Lu, F., Case, B., & Warner, C. M. (2012). Self-disclosure and Mental Health Service use in socially anxious adolescents. *School Mental Health, 4*(4), 219–230. https://doi.org/10.1007/s12310-012-9082-0

de Figueiredo, C. S., Sandre, P. C., Portugal, L. C., Mázala-de-Oliveira, T., da Silva Chagas, L., Raony, Í., Ferreira, E. S., Giestal-de-Araujo, E., dos Santos, A. A., & 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. https://doi.org/10.1016/j.pnpbp.2020.110171

de Lijster, J. M., Dieleman, G. C., Utens, E. M. W. J., Dierckx, B., Wierenga, M., Verhulst, F. C., & Legerstee, J. S. (2018). Social and academic functioning in adolescents with anxiety disorders: A systematic review. *Journal of Affective Disorders, 230*, 108–117. https://doi.org/10.1016/j.jad.2018.01.008

Ebesutani, C., Reise, S. P., Chorpita, B. F., Ale, C., Regan, J., Young, J., Higa-McMillan, C., & Weisz, J. R. (2012). The revised child anxiety and depression scale-short version: Scale reduction via exploratory bifactor modeling of the broad anxiety factor. *Psychological Assessment, 24*(4), 833–845. https://doi.org/10.1037/a0027283

Fröjd, S. A., Nissinen, E. S., Pelkonen, M. U. I., Marttunen, M. J., Koivisto, A., & Kaltiala-Heino, R. (2007). Depression and school performance in middle adolescent boys and girls. *Journal of Adolescence, 31*(4), 485–498. https://doi.org/10.1016/j.adolescence.2007.08.006

Lawrence, D., Dawson, V., Houghton, S., Goodsell, B., & Sawyer, M. G. (2019). Impact of mental disorders on attendance at school. Australian Journal of Education, 63(1), 5–21. https://doi.org/10.1177/0004944118823576

Li, Y., & Lerner, R. M. (2011). Trajectories of school engagement during adolescence: Implications for grades, depression, delinquency, and substance use. *Developmental Psychology, 47*(1), 233–247. https://doi.org/10.1037/a0021307

Lo, W. (2023). *Westborough High School mental health predictive analyses* [Major qualifying project, Worcester Polytechnic Institute]. Digital WPI. https://digital.wpi.edu/show/p5547w00p

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., & Simoes-e-Silva, A. C. (2020). How is covid-19 pandemic impacting mental health of children and adolescents? *International Journal of Disaster Risk Reduction, 51*, 101845. https://doi.org/10.1016/j.ijdrr.2020.101845

Millings, A., Buck, R., Montgomery, A., Spears, M., & Stallard, P. (2012). School connectedness, Peer Attachment, and self-esteem as predictors of adolescent depression. *Journal of Adolescence, 35*(4), 1061–1067. https://doi.org/10.1016/j.adolescence.2012.02.015

Mossman, S. A., Luft, M. J., Schroeder, H. K., Varney, S. T., Fleck, D. E., Barzman, D. H., Gilman, R., DelBello, M. P., & Strawn, J. R. (2017). The Generalized Anxiety Disorder 7-item scale in adolescents with generalized anxiety disorder: Signal detection and validation. *Annals of Clinical Psychiatry : Official Journal of the American Academy of Clinical Psychiatrists, 29*(4), 227–234.

Nguyen, T.-T., Huang, J. Z., & Nguyen, T. T. (2015). Unbiased feature selection in learning random forests for high-dimensional data. *The Scientific World Journal*, 2015, 1–18. https://doi.org/10.1155/2015/471371

Pascoe, M. C., Hetrick, S. E., & Parker, A. G. (2019). The impact of stress on students in Secondary School and Higher Education. *International Journal of Adolescence and Youth, 25*(1), 104–112. https://doi.org/10.1080/02673843.2019.1596823

Richardson, L. P., McCauley, E., Grossman, D. C., McCarty, C. A., Richards, J., Russo, J. E., Rockhill, C., & Katon, W. (2010). Evaluation of the patient health questionnaire-9 item for detecting major depression among adolescents. *Pediatrics, 126*(6), 1117–1123. https://doi.org/10.1542/peds.2010-0852

Stirling, K., Toumbourou, J. W., & Rowland, B. (2015). Community factors influencing child and adolescent depression: A systematic review and meta-analysis. *Australian & New Zealand Journal of Psychiatry, 49*(10), 869–886. https://doi.org/10.1177/0004867415603129

Appendix

	Gender		Race/Ethnicity		All (n = 520)
	F (n = 248)	M (n = 271)	A (n = 174)	W (n = 308)	
Depression Risk	12.10% (30)	6.27% (17)	4.60% (8)	10.39% (32)	9.23% (48)
Anxiety Risk	8.87% (22)	1.85% (5)	1.72% (3)	5.84% (18)	5.19% (27)
Endorse Q18	35.48% (88)	28.04% (76)	31.03% (54)	31.82% (98)	31.73% (165)

"Positive" is a T-score > 65 (corresponds to borderline clinical and clinical) for RCADS Anxiety and Depression scales, and any non-zero response to Q18 ("Think about death")

Table 7 - Risk Distribution in Middle School Students [Positive Rate (N Positive)]

	Gender		Race/Ethnicity		All (n = 489)
	F (n = 267)	M (n = 221)	A (n = 166)	W (n = 283)	
Depression Risk	19.85% (53)	8.14% (18)	10.24% (17)	17.31% (49)	14.72% (72)
Anxiety Risk	7.87% (21)	16.74% (37)	12.65% (21)	11.31% (32)	11.86% (58)
Endorse Q9	11.24% (30)	6.79% (15)	7.83% (13)	9.19% (26)	9.20% (45)

"Positive" is a total score > 9 for the PHQ-9 (corresponds to moderate, moderately severe, and severe) and GAD-7 (corresponds to moderate and severe), and any non-zero response to Q18 ("Thoughts that you would be better off dead")

Table 8 - Risk Distribution in High School Students [Positive Rate (N Positive)]

	F1	Accuracy	Balanced Accuracy
Depression Risk Binary	0.820	0.938	0.836
Anxiety Risk Binary	0.632	0.935	0.627
Endorse Q18	0.695	0.731	0.700

Table 9 -RFC Performances in All Middle School Students

	F1	Accuracy	Balanced Accuracy
Depression Risk Binary	0.765	0.888	0.751
Anxiety Risk Binary	0.625	0.892	0.613
Endorse Q9	0.669	0.908	0.660

Table 10 - RFC Performances in All High School Students

	F1	Accuracy	Balanced Accuracy
All (520)	0.820	0.938	0.836
Female (248)	0.787	0.912	0.792
Male (271)	0.642	0.937	0.663
Asian (174)	0.633	0.937	0.638
White (308)	0.882	0.955	0.894

Table 11 - RFC Performance for Depression Risk Prediction Across Gender and Race/Ethnicity Subgroups of Middle School Students (N Students in Subgroup)

	F1	Accuracy	Balanced Accuracy
All (520)	0.632	0.935	0.627
Female (248)	0.688	0.919	0.691
Male (271)	0.693	0.974	0.744
Asian (174)	0.846	0.983	0.850
White (308)	0.612	0.938	0.615

Table 12 - RFC Performance for Anxiety Risk Prediction Across Gender and Race/Ethnicity Subgroups of Middle School Students (N Students in Subgroup)

	F1	Accuracy	Balanced Accuracy
All (520)	0.695	0.731	0.700
Female (248)	0.666	0.694	0.672
Male (271)	0.667	0.741	0.671
Asian (174)	0.699	0.749	0.712
White (308)	0.671	0.714	0.673

Table 13 - RFC Performance for Endorse Q18 ("Think about death") Prediction Across Gender and Race/Ethnicity Subgroups of Middle School Students (N Students in Subgroup)

	F1	Accuracy	Balanced Accuracy
All (489)	0.765	0.888	0.751
Female (267)	0.782	0.873	0.774
Male (221)	0.547	0.896	0.555
Asian (166)	0.660	0.922	0.665
White (283)	0.736	0.859	0.721

Table 14 - RFC Performance for Depression Risk Prediction Across Gender and Race/Ethnicity Subgroups of High School Students (N Students in Subgroup)

	F1	Accuracy	Balanced Accuracy
All (489)	0.625	0.892	0.613
Female (267)	0.531	0.914	0.533
Male (221)	0.628	0.838	0.635
Asian (166)	0.537	0.861	0.558
White (283)	0.598	0.884	0.592

Table 15 - RFC Performance for Anxiety Risk Prediction Across Gender and Race/Ethnicity Subgroups of High School Students (N Students in Subgroup)

	F1	Accuracy	Balanced Accuracy
All (489)	0.669	0.908	0.660
Female (267)	0.728	0.903	0.698
Male (221)	0.590	0.932	0.593
Asian (166)	0.579	0.921	0.590
White (283)	0.726	0.930	0.713

Table 16 - RFC Performance for Endorse Q9 ("Thoughts that you would be better off dead") Prediction Across Gender and Race/Ethnicity Subgroups of High School Students (N Students in Subgroup)