A Machine Learning Approach to Developing an Adolescent Depression Risk Calculator

Royi Rozen (rr789)
Cornell Tech
Connective Media
rr789@cornell.edu

Parsa Nilchian (Pn292)
Weill Cornell Medicine
Department of Neuroscience
pan4004@med.cornell.edu

Louisa Schilling (lss256)
Weill Cornell Medicine
Department of Neuroscience
lss4002@med.cornell.edu

1. abstract

Background: In the U.S., one in five adolescents experience depression. However, clinical diagnosis of major depressive disorder (MDD) during adolescence is challenging due to the heterogeneity of symptom expression. Moreover, validated tools for assessing depression risk based on behavioral features during early adolescence are not yet widely accessible, particularly for use by parents or guardians.

Purpose: This project aims to develop a depression risk calculator that leverages easily accessible, data-driven predictors of depressive symptoms during early adolescence to identify adolescents at elevated risk for developing symptoms in later adolescence. We employ supervised machine learning to support early risk stratification based on non-invasive baseline behavioral and environmental features.

Methods: We used data from the Adolescent Brain Cognitive Development (ABCD) study, a large-scale longitudinal dataset. Specifically, we analyzed behavioral, environmental, and physical health data collected at baseline (N = 10,0094; ages 9–10) to predict depression outcomes at a three-year follow-up. We framed depression risk as a binary classification problem, using a clinician-inspired outcome variable at three-year follow-up. Two supervised learning methods were applied: logistic regression with LASSO regularization and random forest classification. We initially built our models with all 1,560 features we had access to. We then used our model output (e.g. regressors from logistic regression to identify the most informative predictors. To build a clinically applicable, user-friendly tool, we extracted the top 12 features from each model and re-trained each model using only those variables. Model performance was evaluated using recall, accuracy, and precision, as well as ROC-AUC curves. Regressors from the best-performing model were transformed into a user-friendly risk calculator to stratify adolescents into high- and low-risk groups.

Results: Key predictive features in the logistic regression model included variables describing appetite, sleep, family conflict, family history of mental illness, and participation in sports. The random forest model identified social behaviors, meeting developmental milestones, and perceived extent of social interactions as top predictors. When comparing models trained on their respective top 12 features, the LASSO logistic regression model demonstrated superior recall in identifying adolescents who later developed depressive symptoms (logistic = 0.83; RF = 0.81), accuracy (logistic = 0.62; RF = 0.53) and precision (logistic = 0.47; RF = 0.42). Due to the fact that we intended to build a screening too for parents, we chose the logistic regression as the more suited model for our calculator, as this model had a higher recall (lower risk of missing at risk patients).

Significance: This project translates machine learning findings into a practical, accessible tool for early identification of adolescent depression risk. It supports real-world prevention by using easily obtainable, non-clinical data to inform risk stratification on a vulnerable group of patients.

2. Introduction

2.1. Problem at hand

Depression often first emerges during adolescence, and it is linked to severe long-term outcomes, including substance use, suicidality, and co-morbid psychiatric conditions, and lower economic productivity [3, 4, 6, 14]. Early identification of depression has major benefits as medication and psychotherapy are effective treatment. However, identifying depressive risks in adolescence remains a challenge due to the heterogeneity of symptom expression and risk factors. Current clinical tools — such as symptom checklists — fail to integrate multi-dimensional risk data, limiting their predictive power and real-world applicability. Furthermore, most validated clinical tools, such as the Hamilton Depression Rating Scale (HAM-D) and Montgomery–Åsberg Depression Rating Scale (MADRS) have been only developed for and validated in adults [16].

2.2. Prior efforts

The majority of machine learning research that leverages multi-dimensional data to make predictions on outcomes focuses on hard-to-access or invasive data such as neuroimaging, genetic, bio-specimen, or neuro-cognitive tests [19]. To our knowledge, none of the work on adolescent depression risk has been translated into a user-friendly tool for parent-use.

First, supported by recent work [7], we use logistic regression with LASSO regularization for feature selection due to its interpretability and ability to handle high-dimensional data, making it well-suited for identifying sparse, clinically relevant predictors. Second, we apply a random forest classifier to capture non-linear interactions and obtain robust variable importance rankings, as supported by previous work [21, 2]. What makes our approach unique is its combination of data from high-dimensional data from diverse domains and the translation of our findings into a deployable tool to provide immediate, actionable risk feedback. Unlike prior studies that stop at prediction, our project bridges prediction and implementation, supporting early screening efforts in real-world settings.

While tools such as the Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS) and Child Behavior Checklist (CBCL) assess depressive symptoms in youth, they do not incorporate multi-dimensional risk factors [1, 15]. Moreover, commonly used clinician-administered scales — such as the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V), Hamilton Depression Rating Scale (HAM-D), and Montgomery–Åsberg Depression Rating Scale (MADRS) — lack full validation in adolescent populations and vary in applicability across clinical settings [20, 10]. Recent machine learning studies have aimed to predict depression risk in adolescents, but most rely on narrow feature sets (e.g., neuroimaging or behavioral data alone) [11]. While some work has incorporated multi-domain risk factors, these findings have not been translated into actionable risk estimates or implemented in practical tools for clinical or community use [12]. In contrast, our approach integrates diverse predictors spanning psychological, behavioral, environmental, and lifestyle domains to inform a deployable, user-friendly application for early depression risk assessment in adolescents.

2.3. Summary of ML approach

We propose a supervised machine learning approach to identify early predictors of adolescent depression from the ABCD Study dataset. All data for our models was drawn from the baseline behavioral, environmental, and physical health data collected at age 9 years. After initial data exploration using descriptive statistics and visualizations (e.g., histograms, box plots, scatterplots), we performed rigorous pre-processing (outlier removal, mean/mode imputation for missingness, one-hot encoding for categorical variables, standardization using z-scoring, correlation-based feature reduction). We then performed a binary classification, predicting our outcome variable defined as high-risk or low risk depressive symptoms at the 3-year follow-up (1 = high-risk, 0 = low-risk). This outcome variable was developed based on the Child Behavior Checklist (CBCL) checklist collected at 3-year follow-up. The variable was binarized as per clinical guidance and the top 25th percentile was classified as high-risk.

We used two models: (1) logistic regression with LASSO, and (2) random forest classifiers. Models were trained and validated using a 70/15/15 data split with stratified sampling and 5-fold cross-validation. We addressed class imbalance through stratified resampling and assess performance using ROC curves, confusion matrices, accuracy, precision, and recall/sensitivity. We chose these metrics because they are easy to understand and complement each other. Sensitivity in particular was important for our model since we wanted to build a screening tool that does not miss any adolescents that are at risk (low false negative rate).

We chose logistic regression with LASSO because this is a simple linear model suited to perform binary classification task. The advantage of adding LASSO regularization to the method is that the weights of unimportant features can drift towards zero, allowing us to remove uninformative features from the model and select only highly predictive ones for the final model. Random forest is a good complement to our logistic regression because it can model non-linear relationships and is computationally simple to perform.

After selecting our logistic regression model for its superior recall/sensitivity, we deployed the model in a Streamlit application. Users input responses to questions derived from twelve top features of the model, and the app return a personalized depression risk profile with accompanying visualizations, recommendations, and disclaimers.

2.4. Broader impact

Adolescent depression is a leading contributor to global disease burden and often goes undetected until symptoms escalate. By integrating behavioral, environmental, and psychosocial predictors from a large, longitudinal dataset, our application enables early identification of at-risk youth, potentially guiding timely interventions. However, predictive modeling in mental health raises critical ethical considerations. These include risks of stigmatization, misclassification, and the potential misuse of sensitive information. To mitigate these risks, our model will prioritize interpretability, undergo fairness evaluations, and include clear documentation of intended use. The tool will not be positioned as a diagnostic substitute but rather as a supplementary screening aid to inform clinical judgment. By enhancing transparency and responsible design, our project aims to advance equitable, ethical use of machine learning in adolescent mental health.

3. Background

3.1. Project Scope

Adolescent depression is a growing public health concern, with early onset often predicting worse outcomes in adulthood [18]. Despite affecting almost one in eight teenagers, diagnosis remains delayed or missed in many youths due to barriers such as limited access to specialists, stigma, and reliance on time-consuming or invasive assessments (e.g., neuroimaging or genetic testing) [8]. To address this gap, we built a data-driven, accessible screening tool for elevated depression risk in adolescents.

3.2. Comparisons to prior work

Most prior machine learning research in adolescent mental health has focused on high-dimensional but hard-to-access data such as neuroimaging, genetic markers, or biospecimens [19]. For example, early studies used fMRI data to predict depression onset in adolescents, showing promising results but requiring expensive and specialized imaging resources [9]. Subsequent research incorporated neurocognitive and clinical interview data, improving prediction but still lacking scalability [13].

3.3. Common knowledge

There is widespread agreement among clinicians and researchers that early identification of adolescent depression is critical to improving long-term mental health outcomes. However, there is ongoing debate about the best approach to screening. Some experts advocate for advanced biomarkers such as neuroimaging or genetic data, which may offer biological insights but are expensive, invasive, and impractical for widespread use. Others emphasize self-report measures, which are accessible but criticized for being subjective and less precise.

Organizations like the American Academy of Pediatrics and the U.S. Preventive Services Task Force now recommend regular depression screening during adolescence [17], yet there is no consensus on which specific tools or data types should be used. Researchers also disagree on whether machine learning models should prioritize interpretability (e.g., logistic regression) or predictive power (e.g., black-box models like neural networks).

3.4. Knowledge gaps

Previous research on adolescent depression prediction has often relied on expensive or invasive data modalities such as neuroimaging, genetic testing, or in-depth clinical interviews. While these approaches can yield high predictive accuracy, they are impractical for widespread use in community or school settings. Furthermore, no tool has been directly developed for parents to assess their child's depression risk. In contrast, our study addresses this gap by developing models using only non-invasive, survey-based data, which are far more scalable and accessible. We also apply manual logistic regression with LASSO feature selection, allowing us to identify a concise and interpretable set of predictive variables. This departs from prior work that often uses large, opaque feature sets with limited transparency. Additionally, our tool was designed for parents, which act as their children's healthcare proxies in many instances, as an adolescent affected by depression might not have the insight to seek medical care on their own. By focusing on both interpretability and real-world utility, our approach fills a critical gap between research-oriented modeling and clinical deployment.

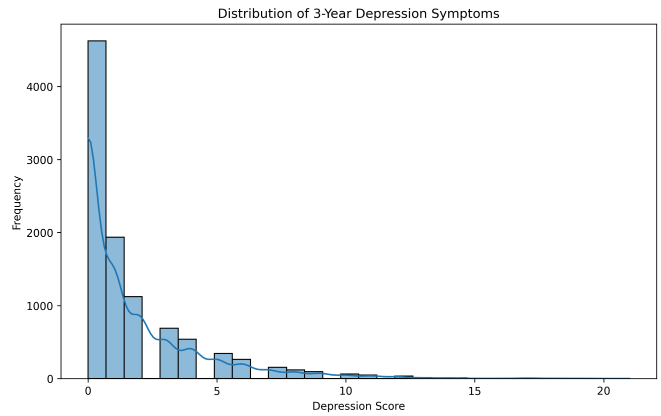

Figure 1: Histogram of depression symptom scores (continuous) in the cohort at the three-year follow-up visit. We then binarized this score (top quartile versus rest) as used as our target variable to predict.

4. End-to-End ML Pipeline

4.1. Back-End

4.1.1 Data Collection, Exploration & Processing

We leveraged high-dimensional data from the Adolescent Brain Cognitive Development (ABCD) Study—a large, longitudinal dataset that follows children beginning at ages 9 to 10 over a ten-year span of development [5]. Data from the ABCD Study is publicly available through the National Institutes of Mental Health (NIMH) Data Archive at: https://nda.nih.gov/abcd. This rich longitudinal design allowed us to use baseline data to predict depressive symptomatology measured at the three-year follow-up visit, providing reliable outcome labels for whether children developed elevated depression symptoms. Baseline data were first released in 2018, with the three-year follow-up data released in 2021, and our project used data up to the 5.1 release in 2023.

Although the ABCD Study includes a wide array of data types—such as genetic, neuroimaging, neurocognitive assessments, biospecimens, and child self-report—for the purposes of this project we focused specifically on parent-reported measures. This decision was motivated by our aim to ensure real-world applicability by using data that could feasibly be known or easily collected by parents or caregivers. These parent-report features spanned a diverse range of domains, including demographics, family structure and functioning, substance use history, physical and mental health, cultural and environmental context, and technology use. The dataset included numerical, ordinal, and categorical variables.

From the full ABCD cohort of 11,868 children with baseline data, 10,094 participants had complete parent-reported depression symptom data at the three-year follow-up. We restricted our analysis to this subset. Depressive symptoms were assessed using parent responses to the Childhood Behavior Checklist (CBCL), from which we derived a depression symptom score (range: 0–20), based on DSM-V-aligned items. Following expert clinical guiance, we binarized this score: children in the top quartile were labeled as high risk (1), and all others as low risk (0). This resulted in N = 7,862 subjects with low depression and 2,412 subjects with high depression. See Figure 1 for the distribution of the raw depression scores.

To construct a meaningful and manageable feature set aligned with our project's goals, we first selected variables from sub-domains relevant to child depression risk: general/demographic, substance use, mental health, sex/gender identity, physical health, and environment/culture. We retained only variables that met the following criteria:

- At least 75% complete across the 10,094 included subjects;
- Sufficient variance (i.e., less than 95% of identical responses);
- Non-redundant (e.g., removed duplicate instances of age, sex, etc.);
- Accessible and interpretable (e.g., excluded text responses, timestamps, and biospecimen measures);
- Encoded as numeric or categorical data.

We then implemented a rigorous pre-processing pipeline, justified as follows:

- **Imputation:** To handle missingness while preserving sample size and avoiding listwise deletion bias, we applied mean imputation for continuous/ordinal variables and mode imputation for categorical/binary variables. These methods are standard for large-scale, survey-based datasets with relatively low missingness.
- **Exploratory Data Analysis:** We used histograms, boxplots, scatterplots, and summary statistics (mean, median, mode, standard deviation, interquartile range, and number of unique values) to examine distributional properties and detect irregularities. This allowed us to identify variables requiring transformation or exclusion. See Figures 2 for examples of a continuous variable.
- **Outlier Removal:** We excluded extreme values defined as beyond ± 1.5 times the interquartile range (IQR). This is a widely used, robust method that reduces the influence of non-representative outliers while maintaining the majority of valid observations.
- **Encoding:** We one-hot encoded categorical variables to allow compatibility with machine learning algorithms while preserving category identity. Ordinal and continuous features were standardized via z-scoring to place all predictors on a comparable scale—an important step for regularized regression models and many distance-based classifiers.
- **Colinearity:** To reduce redundancy and improve model interpretability, we removed highly collinear features based on pairwise correlation thresholds $|r| > 0.7$. Of collinear features, we kept the one that was found to have the greatest absolute value of the correlation with 3-year follow up depression score (continuous, not binary). In all, 31 variables were removed for collinearity.

This resulted in 1,560 baseline features. For both types of machine learning methods, we performed dimensionality reduction by building two stages of models: one full (all features) and one with just the top 12 predictors.

4.1.2 Methods and Model Training

We used two supervised machine learning algorithms to predict the risk of adolescent depression at a 3-year follow-up from baseline data collected at age 10 years: Logistic Regression with LASSO regularization and Random Forest Classification. We chose supervised learning because our task is a classification task (depression vs. no depression) with a known (labeled) outcome variable. We framed the problem as a binary classification task. As per clinical guidance we predicted whether a child falls in the top 25% of depression severity scores 3 years later (binary outcome: 1 = high risk/depression, 0 = low risk/no depression).

1. Logistic Regression with LASSO regularization: A logistic regression with LASSO correction was performed

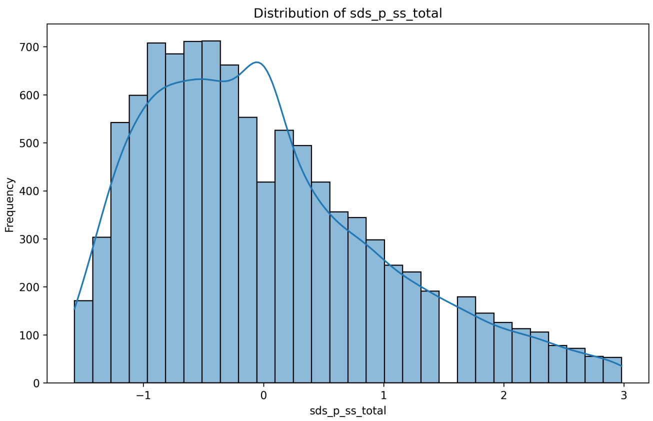

Figure 2: Histogram showing distrubution of continuous variable `sds_p_ss_total` or "Social behavior and communication" after z-scoring.

to identify the variables with the highest predictive power for the outcome of depression. This method has the advantage of being easy to interpret but the disadvantage is that it only captures linear relationships. We obtained a coefficient for each feature and then used these to estimate probabilities of depression risk. A recent study found logistic regression was comparable to other ML models in predicting depression and identifying risk factors while also exhibiting less risk of overfitting [**?**].

Algorithmically, the logistic regression model was trained using gradient descent to minimize the cross-entropy loss between predicted and actual binary labels. The loss function is defined as

$$\mathcal{L}(\mathbf{w}) = -\frac{1}{n}\sum_{i=1}^{n}\left[y^{(i)}\log\left(h^{(i)}+\varepsilon\right)+(1-y^{(i)})\log\left(1-h^{(i)}+\varepsilon\right)\right] \tag{1}$$

Predictions were generated using the sigmoid function applied to a weighted sum of standardized input features. The model was trained with a learning rate of 0.01 over 5000 iterations, which allowed us to empirically monitor convergence and loss minimization. After training, the model outputted a probability score for each individual, and we applied a custom threshold of 0.25 (instead of the conventional 0.5) to classify individuals into high- or low-risk groups, thereby increasing sensitivity and reducing false negatives.

2. Random Forest Classification: We built an ensemble of decision trees trained on bootstrapped samples from the training data. Gini impurity was used as the node-splitting criterion, defined as:

$$\text{Gini}(t) = 1 - \sum_{i=1}^{C} p_i^2 \tag{2}$$

where p_i is the proportion of samples belonging to class i at node t, and C is the number of classes. Probability of depressive risk mas made by proportion of trees voting for each class and prediction was made by majority vote across the ensemble of decision trees.

The advantage of random forest is that it captures nonlinear relationships and provides interpretable feature importance rankings. However, interpretation of individual trees can be challenging. Prior studies have found that random forest classifiers are effective for predicting depression in youth, often outperforming other machine learning methods [21, 2].

Inputs: First - for each machine learning method method - we created fulls models including 1,560 number preprocessed baseline features including variables assessing demographics (sex, gender, race, age, puberty status), behavior (executive functioning, emotional reactivity), environmental (school, family, and social), behavior (sleep, screen time, physical activity), family history of mental illness, and early life adversity factors. Second - for each machine learning method - we created another model using just our subset of 12 top predictive features (determined in the full model) as inputs.

Outputs: Prediction of depressive symptomatology at follow-up assessed via binarization of parent responses to the CBCL questionnaires at ages 13-14.

4.1.3 Model Evaluation

Evaluation metrics: We prioritized recall (sensitivity) in our model, as the primary goal of a risk assessment tool is to identify as many at-risk individuals as possible. In the context of predicting whether a child will develop depression, it is preferable to adopt a more lenient threshold that errs on the side of inclusion—minimizing false negatives — so that children who may be at elevated risk are not overlooked. This approach aligns with the preventive nature of youth mental health screening, where early identification can facilitate timely support and intervention during critical developmental periods, even if it results in a higher false positive rate.

Nonetheless, other evaluation metrics such as accuracy, precision, area under the ROC curve (AUC), and F1 score were important for us to comprehensively assess our model's performance. Accuracy provides an overall sense of correctness, though it may be less informative in this context due to the class imbalance in our outcome variable (75/25 split). Precision quantifies the proportion of predicted high-risk cases that are truly at risk, offering insight into the burden of false positives—which is relevant when considering the potential for unnecessary concern or misallocation of clinical resources. AUC measures the model's ability to distinguish between at-risk and not-at-risk children across all thresholds, offering a threshold-independent view of discriminatory power. F1 score, the harmonic mean of precision and recall, offers a balanced metric that is especially useful when both false positives and false negatives carry meaningful consequences. Together, these metrics provide a more nuanced understanding of our model's strengths and limitations in guiding early depression risk assessment.

Train-Validation-Test Split: Following pre-processing, we split the dataset into training (70%), validation (15%), and test (15%) subsets. We also accounted for class imbalance in the binary outcome during model training and evaluation by stratified sampling such that there was a 75/25 split of non-depressed/depressed subjects in each set. This was done for both the full model (all features) and limited model (12 top features) for both logistic regression and random forest.

Hyperparameter Optimization:

We performed empirical hyperparameter tuning by observing the loss over iterations during training. By experimenting with different learning rates and iteration counts, we selected values that produced stable convergence and minimized the cross-entropy loss without overfitting.

Random forest models were trained on the training set (70% of data) for each combination of hyperparameters and then tested on the validation set (15%). Finally - after using a model built on the training dataset with the hyperparamters that yielded an optimal AUC score - the final model was tested on the held-out test data (15%) and we report the performance of this here. Class balance was held constant across all three sets (25% depressed). Optimal hyperparameters (i.e., number of trees, maximum depth, and minimum sample split) were determined using a grid search on the validation set based on F1 score optimization. Number of trees were optimized across values 50, 100, and 200. Max depth was optimized across 5, 10, and 15. Minimum sample split was optimized across 2, 5, and 10. Feature importance was calculated as the total Gini impurity reduction attributed to each feature across all trees in the forest. The top 12 features were selected to build a simplified model using the same a 70-15-15 split and the same hyperparamter optimization procedure.

Avoiding Overfitting/Underfitting: For both logistic regression and random forest models, we used a stratified train/validation/test split, keeping the test set completely untouched during model selection. This approach prevented data leakage and ensured an unbiased estimate of generalization performance. Additionally, re-running both models using only the top 12 most important features reduced input dimensionality, thereby lowering the risk of overfitting.

In the random forest model, we tuned key hyperparameters —maximum depth, minimum samples per split, and number of trees — across a defined grid (explained above) to control model complexity and prevent individual trees from overfitting. Optimizing hyperparameters based on AUC, a threshold-independent metric, helped avoid selecting a model that performed well on training data but poorly on unseen data.

Random forests are inherently capable of modeling complex non-linear relationships, making them less prone to underfitting than logistic regression. To further address underfitting, our hyperparameter grid included values that allowed for greater model flexibility when needed. Moreover, we initially included a large number of features in the full model before reducing to the top 12, which allowed the model to capture potentially informative signals and further reduced the risk of underfitting.

To prevent overfitting, we applied LASSO regularization, which penalizes large coefficients and encourages sparsity—automatically shrinking less informative variables toward zero. This promotes a simpler model that generalizes better to new data. Additionally, we manually restricted the model to the top 12 most predictive features, based on coefficient magnitude, further reducing model complexity and risk of overfitting. We monitored training loss and evaluated performance on a held-out validation set to ensure the model was neither underfitting nor overfitting.

4.1.4 Results

Logistic regression:

The most influential predictors, identified by the largest magnitude coefficients for the logistic regression, included `ksads_sleepprob_raw_814_p` ("Parent-reported sleep problems"), `cbcl_q71_p` ("Self-conscious or easily embarrassed"), `cbcl_q04_p` ("Cries a lot"), `famhx_ss_parent_prf_p` ("Parental history of psychiatric problems"), `sds_p_ss_does` ("Social engagement and actions"), `cbcl_q86_p` ("Feels too guilty"), `cbcl_q09_p` ("Fears doing bad things"), `asr_q59_p` ("Adult somatic complaints"), `asr_q47_p` ("Adult anxiety symptoms"), `cbcl_q112_p` ("Worries"), `sds_p_ss_total` ("Social behavior and communication"), and `cbcl_q22_p` ("Feels worthless or inferior"). These variables reflect emotional dysregulation, sleep disturbances, family psychiatric history, and social functioning—all of which align with known risk factors for adolescent depression. See Figures 3 and 4 for the ROC curve and confusion matrix of the logistic regression model of the top 12 features.

Logistic Regression (full model):

- Recall/Sensitivity: 0.82
- AUC: 0.73

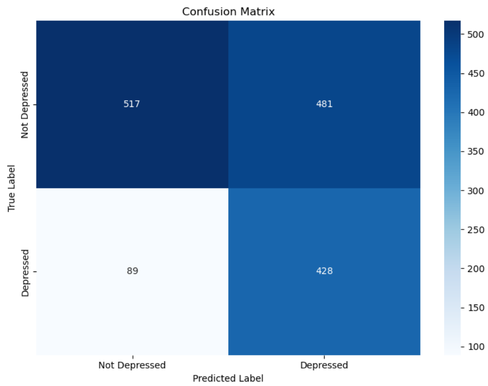

Figure 3: Confusion matrix for logistic regression with LASSO (top 12 predictors only).

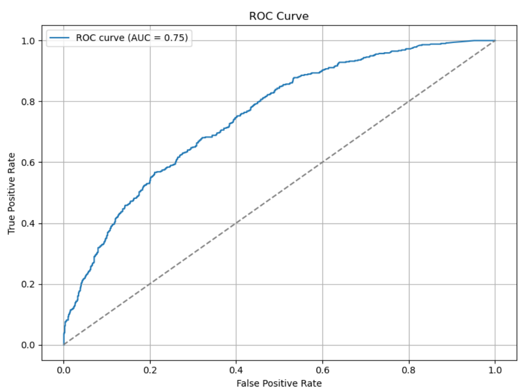

Figure 4: A receiver operating characteristic curve (ROC) for logistic regression with LASSO (top 12 predictors only).

- Accuracy: 0.61
- Precision: 0.47

Logistic Regression (top 12 features):

- Recall/Sensitivity: 0.83
- AUC: 0.75
- Accuracy: 0.62
- Precision: 0.47

Random Forest:

The random forest classification model using all features yielded decent performance metrics. For the full model, the optimal hyperparameters were: `n_trees=200`, `max_depth=15`, `min_samples_split=10`. The top predictors based on total Gini impurity reduction included `sds_p_ss_total` ("Social behavior and communication"), `devhx_18_p` ("Prenatal substance exposure"), `devhx_3_p` ("Early developmental milestones"), `demo_prnt_years_us_v2` ("Years parent has lived in the US"), `demo_prnt_income_v2` ("Parental income"), `meim_p_ss_total` ("Ethnic identity strength"), `macv_p_ss_fr` ("Friendship support subscale"), `macv_p_ss_fs` ("Family support subscale"), `sai_ss_lmusic_hours_p` ("Hours per week playing music"), `sds_p_ss_dims` ("Social responsiveness"), `sds_p_ss_does` ("Social engagement and actions"), and `sai_ss_lmusic_years_p` ("Years of music lessons").

Random Forest Classification (full model):

- Recall/Sensitivity: 0.75
- AUC: 0.75
- Accuracy: 0.66
- Precision: 0.51

Random Forest Classification (top 12 features):

- Recall/Sensitivity: 0.81
- AUC: 0.67
- Accuracy: 0.53
- Precision: 0.42

We selected the Logistic Regression with LASSO model for deployment based on critical considerations for our depression risk screening tool. There was superior recall, accuracy, and precision for logistic regression compared to random forest as it represents the model's ability to correctly identify children who are actually at elevated risk of depression. In a screening context, minimizing false negatives (missed cases of depression risk) takes precedence over minimizing false positives, as the latter can be addressed through subsequent professional assessment. Additionally, the Logistic Regression model offers greater interpretability through its feature coefficients, allowing parents to understand which specific reported behaviors influence the risk estimation, making the insights more actionable and transparent.

4.1.5 Model Deployment

The following section presents a detailed description of the web application developed to deploy our trained machine learning model for adolescent depression risk assessment. This section covers the application development using Streamlit, the model selection process, target population analysis, user interface design, and the technical integration between frontend and backend components.

Web Application Development

The web-based application, titled "Child Behavioral Insights Estimator," was developed using the Streamlit Python library to provide an accessible and user-friendly tool for parents to gain preliminary estimations of their child's behavioral patterns associated with depression risk. The application functions as a translation layer, converting complex machine learning insights derived from the Adolescent Brain Cognitive Development (ABCD) Study into understandable outputs for non-technical users (see figures 5 and 6. A critical design element of the application is its clear presentation as an estimation tool rather than a diagnostic instrument, with prominent disclaimers advising users to consult qualified healthcare professionals for definitive assessments or concerns. This careful positioning supports responsible use of machine learning in mental health contexts while still providing valuable risk stratification capabilities. The deployment phase required careful evaluation of trade-offs between our two trained models to determine which would best serve the application's purpose. We analyzed performance metrics from both models to inform this decision:

Target Population and Benefits

The primary target population for our application is parents or guardians of adolescents. More specifically, the tool caters to parents who are seeking accessible preliminary assessments of their child's behavioral risks related to depression but may not have immediate access to clinical services or professional guidance. This population often wishes to be more informed about potential mental health concerns before pursuing formal evaluation.

The benefits to this population are multifaceted. First, the application empowers parents with data-informed perspectives on their child's behavioral patterns, facilitating early awareness of potential issues. By providing early risk estimation, the tool may prompt parents to seek professional advice or implement supportive measures sooner, potentially leading to improved intervention outcomes. The web-based format offers a convenient, private, and no-cost method for parents to explore concerns without the immediate need for clinical visits, lowering barriers to information-seeking. Additionally, the interpretability of our chosen Logistic Regression model helps parents understand contributing factors, guiding more informed conversations or observations.

User Interface Layout

The user interface follows an intuitive design prioritizing clarity and ease of use. The application begins with a clear title ("Child Behavioral Insights Estimator") and an introductory section explaining its purpose, data source (ABCD Study), and critical disclaimers emphasizing its non-diagnostic nature. The core input mechanism is an interactive questionnaire about the child's behavior over the past six months, with users providing answers using radio buttons for simple "Yes/No" responses or frequency-based options (e.g., "Never/Sometimes/Often"). This questionnaire directly corresponds to the input features required by our pre-trained logistic regression model.

Figure 5: Start of our website describing the utility of our calculator).

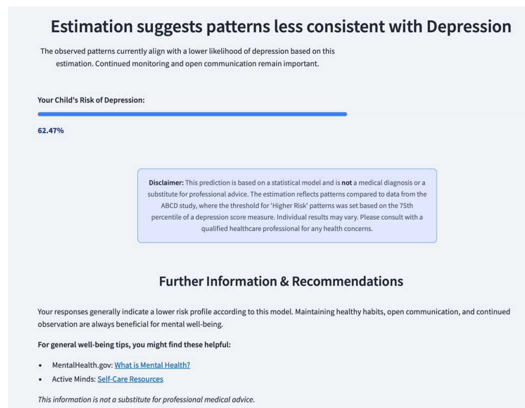

Figure 6: Final output for the user after finishing the questionnaire).

Following questionnaire completion and submission, the interface displays the estimation results in multiple formats. This includes a clear statement of the child's estimated depression risk presented as a percentage (e.g., "55.97%") with appropriate visual representation to enhance comprehension. A prominent disclaimer reiterates that the prediction is based on a statistical model and is not a medical diagnosis. The interface also provides actionable recommendations and links to reputable mental health resources, with control buttons allowing users to toggle section visibility or restart the assessment process.

Frontend-Backend Integration

Streamlit Application Interface showing questionnaire an Estimator interface displaying the questionnaire with radio button inputs and important disclaimers about the non-diagnostic nature of the tool.

Results Display Interface showing risk assessment the calculated depression risk percentage, disclaimers, and resource recommendations for parents seeking additional information.

The frontend and backend components are seamlessly integrated within the Streamlit framework through a well-defined data flow process. When a user interacts with the Streamlit interface and completes the questionnaire, their responses are collected via the radio button inputs and organized into a structured data format. Upon submission, this data is passed to the backend Python functions for processing.

The backend processing begins with the pre-trained Logistic Regression model, which is loaded when the application initializes. User inputs are carefully preprocessed to match the expected format of the model, including appropriate scaling if the model was trained on standardized data (as ours was with z-scoring). Features are properly ordered and formatted into a numerical feature vector suitable for the model's prediction method. The preprocessed input data is then fed into the loaded model's `.predict_proba()` method to calculate the probability of depression risk.

Once the prediction is generated, the backend returns this result to the Streamlit frontend script, which updates the user interface accordingly. The risk percentage is displayed along with visual representations, and conditional logic may present different recommendation texts based on the prediction outcome. This architecture ensures a responsive user experience while leveraging the predictive power of our machine learning model in an accessible format for parents.

4.1.6 Conclusion

We develop a practical, interpretable machine learning model to identify adolescents at elevated risk for depression using non-invasive survey data from the ABCD study. We trained and evaluated two primary supervised learning models: logistic regression with L1 (LASSO) regularization and random forest classification.

Our pipeline involved pre-processing over one thousand survey-derived variables, standardizing features, removing low-variance items, and creating a binary outcome representing the top 25% of 3-year depression scores. LASSO was used for feature selection, reducing the model to the 12 most predictive items across sleep, mood, family history, and psychosocial domains.

Our results have direct value for parents, schools, and pediatric providers who lack access to mental health specialists. By minimizing false negatives and prioritizing recall, our model ensures that at-risk youth are flagged early, potentially facilitating timely interventions. Societally, this supports a scalable, low-cost approach to mental health screening that could reduce the long-term burden of untreated adolescent depression.

In the future, we plan to validate our model's generalizability using held-out ABCD sites or other datasets. We also aim to incorporate user-friendly deployment, such as improving our Streamlit-based risk calculator for parent or clinician use. Lastly, We aim to collaborate with clinical partners to test the tool in real-world screening scenarios and refine its use based on feedback from stakeholders.

5. Team Member Contribution

All team members collaborated on the overall project direction, research, model development, final report, and presentation—including the abstract and conclusion. Individual responsibilities were divided as follows:

Technical Components

- **Louisa Schilling (Data):** Responsible for sourcing the ABCD dataset, cleaning and preprocessing the data, performing exploratory data analysis, and preparing features for modeling.
- **Parsa Nilchian (Machine Learning):** Developed and implemented the machine learning pipeline, including Logistic Regression and Random Forest models. Led model training, evaluation (with emphasis on recall), and final feature selection.
- **Royi Rozen (Streamlit Application & Integration):** Designed and developed the "Child Behavioral Insights Estimator" web application using Streamlit. Handled UI/UX and integrated the trained model for interactive user predictions.

Writing Components

Primary writing responsibilities for the final report were divided as follows:

- **Louisa Schilling:** Drafted sections on data acquisition, preprocessing, and dataset descriptions.
- **Parsa Nilchian:** Wrote the sections on machine learning methodology, training procedures, and evaluation results.
- **Royi Rozen:** Authored content related to the web application, its design, implementation, and deployment process.

All members contributed to writing and refining the introduction, background, risk and mitigation strategies, expected outcomes, and conclusion sections.

References

[1] Thomas M Achenbach, Thomas M Ruffle, et al. The child behavior checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in review*, 21(8):265–271, 2000. 2

[2] Mohammad Hossein Amirhosseini, Adefemi Lawrence Ayodele, and Amin Karami. Prediction of depression severity and personalised risk factors using machine learning on multimodal data. In *2024 IEEE 12th International Conference on Intelligent Systems (IS)*, pages 1–7. IEEE, 2024. 2, 5

[3] Shelli Avenevoli, Joel Swendsen, Jian-Ping He, Marcy Burstein, and Kathleen Ries Merikangas. Major depression in the national comorbidity survey–adolescent supplement: Prevalence, correlates, and treatment. *Journal of the American Academy of Child & Adolescent Psychiatry*, 54(1):37–44, 2015. 1

[4] Joshua Breslau, Elizabeth Miller, W-J Joanie Chung, and Julie B Schweitzer. Childhood and adolescent onset psychiatric disorders, substance use, and failure to graduate high school on time. *Journal of psychiatric research*, 45(3):295–301, 2011. 1

[5] Betty Jo Casey, Tariq Cannonier, May I Conley, Alexandra O Cohen, Deanna M Barch, Mary M Heitzeg, Mary E Soules, Theresa Teslovich, Danielle V Dellarco, Hugh Garavan, et al. The adolescent brain cognitive development (abcd) study: imaging acquisition across 21 sites. *Developmental cognitive neuroscience*, 32:43–54, 2018. 3

[6] William E Copeland, Shari Miller-Johnson, Gordon Keeler, Adrian Angold, and E Jane Costello. Childhood psychiatric disorders and young adult crime: A prospective, population-based study. *American Journal of psychiatry*, 164(11):1668–1675, 2007. 1

[7] Xing-Xuan Dong, Jian-Hua Liu, Tian-Yang Zhang, Chen-Wei Pan, Chun-Hua Zhao, Yi-Bo Wu, and Dan-Dan Chen. Comparison of logistic regression and machine learning approaches in predicting depressive symptoms: A national-based study. *Psychiatry Investigation*, 22(3):267, 2025. 2

[8] Renee D Goodwin, Lisa C Dierker, Melody Wu, Sandro Galea, Christina W Hoven, and Andrea H Weinberger. Trends in us depression prevalence from 2015 to 2020: the widening treatment gap. *American Journal of Preventive Medicine*, 63(5):726–733, 2022. 3

[9] Zeus Gracia-Tabuenca, Elise B Barbeau, Yu Xia, and Xiaoqian Chai. Predicting depression risk in early adolescence via multimodal brain imaging. *NeuroImage: Clinical*, 42:103604, 2024. 3

[10] Max Hamilton. A rating scale for depression. *Journal of neurology, neurosurgery, and psychiatry*, 23(1):56, 1960. 2

[11] Umme Marzia Haque, Enamul Kabir, and Rasheda Khanam. Detection of child depression using machine learning methods. *PLoS one*, 16(12):e0261131, 2021. 2

[12] Gillian Ho, Danielle N Pratt, Miranda A Bridgwater, Jason Schiffman, Lauren M Ellman, and Vijay A Mittal. Factors impacting intent to seek treatment within youth at clinical high risk for psychosis. *Schizophrenia Research*, 267:273–281, 2024. 2

[13] Tiffany C Ho, Rutvik Shah, Jyoti Mishra, April C May, and Susan F Tapert. Multi-level predictors of depression symptoms in the adolescent brain cognitive development (abcd) study. *Journal of Child Psychology and Psychiatry*, 63(12):1523–1533, 2022. 3

[14] Dylan Johnson, Gabrielle Dupuis, Justin Piche, Zahra Clayborne, and Ian Colman. Adult mental health outcomes of adolescent depression: A systematic review. *Depression and anxiety*, 35(8):700–716, 2018. 1

[15] Joan Kaufman, Boris Birmaher, David Brent, UMA Rao, Cynthia Flynn, Paula Moreci, Douglas Williamson, and Neal Ryan. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (k-sads-pl): initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36(7):980–988, 1997. 2

[16] Stefan Leucht, Hein Fennema, Rolf Engel, Marion Kaspers-Janssen, Peter Lepping, and Armin Szegedi. What does the hamd mean? *Journal of affective disorders*, 148(2-3):243–248, 2013. 1

[17] Carol M Mangione, Michael J Barry, Wanda K Nicholson, Michael Cabana, David Chelmow, Tumaini Rucker Coker, Karina W Davidson, Esa M Davis, Katrina E Donahue, Carlos Roberto Jaén, et al. Screening for depression and suicide risk in children and adolescents: Us preventive services task force recommendation statement. *Jama*, 328(15):1534–1542, 2022. 3

[18] Leslie Miller and John V Campo. Depression in adolescents. *New England Journal of Medicine*, 385(5):445–449, 2021. 3

[19] Ghazal Mirabnahrazam, Da Ma, Sieun Lee, Karteek Popuri, Hyunwoo Lee, Jiguo Cao, Lei Wang, James E Galvin, Mirza Faisal Beg, Alzheimer's Disease Neuroimaging Initiative, et al. Machine learning based multimodal neuroimaging genomics dementia score for predicting future conversion to alzheimer's disease. *Journal of Alzheimer's Disease*, 87(3):1345–1365, 2022. 2, 3

[20] Stuart A Montgomery and MARIE Åsberg. A new depression scale designed to be sensitive to change. *The British journal of psychiatry*, 134(4):382–389, 1979. 2

[21] Radwan Qasrawi, Stephanny Paola Vicuna Polo, Diala Abu Al-Halawa, Sameh Hallaq, Ziad Abdeen, et al. Assessment and prediction of depression and anxiety risk factors in schoolchildren: machine learning techniques performance analysis. *JMIR formative research*, 6(8):e32736, 2022. 2, 5