2019

LONGITUDINAL DATA PREDICTION IN EHR: COMPARISON OF GLMM AND MACHINE LEARNING METHODS

Wenqiu Cao

LONGITUDINAL DATA PREDICTION IN EHR: COMPARISON OF GLMM AND MACHINE LEARNING METHODS

BY

WENQIU CAO

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

2019

MASTER OF SCIENCE THESIS

OF

WENQIU CAO

APPROVED:

Thesis Committee:

Major Professor Theodore A. Walls

Lynda A. R. Stein

Haibo He

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2019

ABSTRACT

In this study, we aimed to develop and compare models to predict individuals with suicidal ideation using Generalized Linear Mixed Model(GLMM) and Machine Learning(ML) algorithms. We conducted secondary data analysis with data collected by an online clinical measurement company. The sample included 402 individuals aged over 18 years who have received more than three psychiatric treatments since 2017. The data were split into a training set(70%) and a testing set(30%) randomly. In the training set, GLMM, RF model, and GBDT model were trained with all the features. Conditional RF and GBDT with variables selected based on GLMM were trained next. Subsequently, the fitted models were used to predict suicide ideation in the test set. All analyses were conducted in *R* and *Python*. The prediction models based on ML algorithms (R^2 from 0.260 to 0.409, MSE from 1.761 to 2.202, MAE from 0.942 to 0.985) performed better than GLMM (R^2 = 0.115, MSE=2.880, MAE=1.013). The insights gained from this study may be of assistance to broadly apply ML algorithms to the massive data from EHR to enhance suicide risk prediction. There is, therefore, a definite need for improvements understanding prediction accuracy versus traditionally employed GLMM approaches.

ACKNOWLEDGMENTS

I appreciate everyone who has taught and supported me throughout my graduate life. I would first like to thank my major professor Dr. Theodore (Ted) Walls. He has been supporting and guiding me in all the time of my graduate study with his patience, motivation, and immense knowledge. I would also like to express my gratitude to the rest members of my committee, Dr. Lynda (Lyn) Stein, Dr. Haibo He, and Dr. Yichi Zhang, for their extraordinary support and insightful comments.

My sincere thanks also go to all of the University of Rhode Island faculty, who instructed me in class. The knowledge and skills they taught me are invaluable. I would also like to thank my labmates: Coral Shuter, Dr. Daniel Smith, Valerie Ryan, and Will Lowenkamp. You supported me a lot and were always willing to help me.

Last but not least, I would like to thank my Mom for giving birth to me and supporting me unfailingly throughout my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Longitudinal Data Prediction in EHR: Comparison of GLMM and Machine Learning Methods

Risk behavior prediction can help to avert disadvantageous outcomes in mental health and health behavior. This determination of antecedent risks to a specific behavior problem or mental disorder can be employed to guide preventative decisions by clinicians, doctors, policymakers, and educators.

With an annual death rate of 800,000 people, suicide has been a leading cause of unnatural death that invites worldwide attention [1]. Arsenault-Lapierre, Kim, and Turecki[2] found in a meta-analysis that most suicide completers have been diagnosed with mental disorders or illnesses, such as depression and substance use disorder. There is evidence that suicidal ideation plays a crucial role in predicting future suicide attempts and behaviors. Therefore, assessing suicidal ideation among individuals having psychiatric treatment is an essential strategy in suicide prevention.

Although risk prediction models are very prevalent in commercially relevant areas such as finance and the physical sciences, mental health data archives have only recently begun to be sufficiently robust for this kind of modeling. Electronic Health Records (EHRs) are commonly used in health care situations, which provide access to longitudinal data that can be used to analyze change over time and predict future outcomes. Machine learning is a branch of artificial intelligence (AI), in which a computer learns from the raw data to generate the underlying rules. The increasing volume of person-specific multivariate trend data in mental health EHRs makes it possible to enhance prediction by utilizing machine learning methods.

The goal of the current research was to compare the viability and performance of suicidal ideation prediction using two approaches on a longitudinal mental health EHR database: a) the Generalized Linear Mixed Model (GLMM) and, b) machine learning algorithms. We expected that machine learning algorithms could provide higher accuracy than GLMM when used to predict individuals at high risk of suicidal ideation.

CHAPTER 2

Literature Review

2.1 Importance of Suicidal Ideation Prediction

Mental disorders are common in the United States. Approximately 18.5% U.S. adults, about 44.7 million, live with a mental disorder, and about 9.8 million, 20% of this group, are suffering severe mental illnesses in 2016 [3]. Within this population, individuals are at an increased risk of involving risk behaviors (e.g., substance use, suicide risk, unsafe sex).

Previous studies reported higher rates of suicidal thoughts and attempts in psychotic populations [4]. Psychosis is found to directly or indirectly contribute to suicide risks. Alcoholism, drug abuse, and self-injurious behaviors are all risky behaviors that individual may use to get temporary relief from intense stress or emotional pain, however, almost always results in greater feelings of loneliness and hopelessness. These types of vicious cycles can cause severe mental illness or can worsen an existing mental health disorder [5], or even threaten one's own life. This underscores the urgency and importance of suicide prevention among psychotherapy patients.

Considered as a significant predictor of committing suicide, the term 'suicidal ideation' was used in this thesis to refer to the thoughts about killing oneself. Even though individuals with suicidal ideation do not all subsequently have suicide behaviors or attempts, patients who have consistent suicidal ideation are at higher risk of committing suicide[6]. Previous research has established that among individuals who have or had suicidal ideation, the probability of future suicide behaviors or attempts was approximately 55%[7]. About 60% of the transitions from suicidal ideation to suicide plan, and then to suicide attempt, occurred within the first year after onset of suicidal ideation[7]. Predicting individuals' levels of suicidal ideation might be an effective approach to target the patients at high risk of suicide and provide relevant interventions.

2.2 Electronic Health Records

Electronic Health Records (EHRs) are real-time, patient-centered records that make health-related information available instantly and securely to authorized users [8]. EHRs replaced paper-based systems in many healthcare organizations and are used commonly to capture and utilize the vast amount of detailed clinical information and offers lots of advantages for clinical research, such as cost efficiency, the considerable amount of data, and the ability to analyze data over time [9]. A recent report of the American Medical Informatics Association stated that [10]:

> Secondary uses of health data can enhance individuals' health care experiences, expand knowledge about diseases and treatments, strengthen understanding of health care system effectiveness and efficiency, support public health and security goals, and aid businesses in meeting customers' needs.

In the United States, due to its broad implementation, accumulated EHR data have become a crucial resource for clinical studies. Notably, access to longitudinal data that can be used to predict future outcomes opens opportunities to support decision making or clinical judgment for patients. Simon et al.[11] used EHRs to develop prediction models for suicide attempt and suicide death over 90 days for both mental health and primary care visits. Their models showed that c-statistics (equivalent to area under the curve) for suicide attempts prediction ranged from 0.833 to 0.861, achieving a significant improvement for the predictive performance than existing prediction tools[11]. EHRs offer a data-rich environment for scientists to conduct essential research and connect with practice in the future.

2.3 Generalized Linear Mixed Model

Longitudinal data refers to the data obtained by observing and recording each participants at successive time points over a period of time. Compared with the cross-sectional data, the main advantage of longitudinal data is that it can more effectively estimate the person-specific and group trends of changes over time within and between samples. Longitudinal data analysis supports modeling of the relations between response variables and covariates, and also considers the covariance of the response variable set. Longitudinal designs are widely used in psychological research. Traditional methods include Latent Growth Curve (LGC), Linear Mixed Model (LMM), Generalized Linear Mixed Model (GLMM), and so on.

The generalized linear mixed model (GLMM) can be used to access predictive relations between selected covariates and random coefficients reflecting the variability of person-specific intercepts or trends [12]. GLMM combines two basic and widely-used statistical methods, the linear mixed model (LMM) and the generalized linear model (GLM). Therefore, it can be used for categorical longitudinal variables and provides a flexible framework for analyzing grouped data while considering the within-group correlation. It can handle non-normal data by using link functions and exponential distributions, involving both fixed and random effects [13].

2.4 Limitations of GLMM

GLMM is a routine option to explore longitudinal prediction for categorical outcomes in behavioral science [14, 15]. However, there are two main limitations while applying GLMM in EHRs to develop predictive models.

First, similar to other linear parametric models, GLMM is predicated on multiple statistical assumptions, including additivity of the linear predictors, in-dependence of errors, equal variance of errors (homoscedasticity) and normality of errors[16][17][18]. Under those assumptions, we can estimate statistical tests and magnitudes according to some criteria. However, these are all based on assumptions about data distribution and models. If these assumptions are not met, the statistical criteria would become meaningless and Type I error rate would increase[17]. Therefore, the traditional method has a strong dependence on the assumptions and theories, which are difficult to verify in some cases.

Second, a challenge in deploying this model is that the use of a large dataset with hundreds of independent variables introduces the possibility of over-fitting [19]. In predictive modeling, we regard the real underlying factors as the *signal*, which we want to learn from the data. *Noise*, conversely, refers to the unnecessary detail or randomness in a specific dataset. When researchers include too much irrelevant information in a regression equation to increase the effect size, the prediction model will be vulnerable to overfitting, resulting in lower suitabilty to be applied to other databases.

2.5 Advantage and Feasibility of Machine Learning Methods

Machine learning is a compelling predictive method for large-scale, high-dimensional data, enabling computers to "learn by themselves" based on data (e.g., progressively improve their performance on a specific task), without the model being directly specified [20]. This method can be used to minimize the problem of overfitting by searching for stable data patterns based on algorithmic rules [21]. The increasing accessibility of big data in health care makes its great potential to enhance health service with data mining and machine learning methods. In addition, common analysis models in psychological studies, such as regression and classification models, can be alternatively pursued via machine learning algorithms to identify linear and nonlinear patterns without predefined underlying assumptions [17]. Identified patterns can be utilized to make predictions about future events and then be continuously used to improve the model performance for better prediction.

Several recent papers have pointed to some success stories when behavioral scientists have employed a predictive approach involving pattern detection [22]. Rosellini et al.[23] developed a classification model based on machine learning technique to predict soldiers at high-risk of violent crime among 975,057 U.S. Army soldiers. Area under curve (AUC) was 0.79 (for both men and women) in the 2004–2009 training set and 0.74–0.82 (men-women) in the 2011–2013 test set. Ding, Bickel and Pan[24] built a social media-based substance use detection systems using unsupervised machine learning and text mining techniques. Walsh, Ribeiro, and Franklin[25] used Random Forest (RF) to target patients attempting suicide which accurately predicted future suicide attempts ($AUC = 0.84$, $precision = 0.79$, $recall = 0.95$, $Brierscore = 0.14$). Notably, Barak-Corren and colleagues[26] used longitudinal EHRs to predict suicide attempt or death among outpatients and achieved sensitive (33%-45% sensitivity), specific (90%-95% specificity), and early (3-4 years in advance on average) prediction of patients' future suicidal behavior.

Therefore, we explored the application of both traditional GLMM and machine learning algorithms to assess the relative performance of modeling strategies in attaining a stable, accurate and efficient suicidal ideation prediction models. The readers should bear in mind that the purpose of the current study was to compare the methods used to build prediction models, rather than to explain or support a theoretical model. We expected this research would provide researchers with some critical guidance on model selection, through a fully worked pair of example involving the same database. A future goal is to standardize our application of machine learning methods on EHR data to inform a real-time data-driven clinical decision support system.

CHAPTER 3

Methodology

3.1 Data

The current study involved secondary data analysis, which has been classified by the University of Rhode Island Institutional Review Board (IRB) as non-human-subject research HU1718-124, as it is a secondary data analysis. The thesis only engaged with part of clinical EHR data. A software company, Mirah[27], collected all information and data, much of which had been formally approved for research purposed at other academic institutions. MIRAH offers routine measurement of patient symptoms to multiple clinic sites all over the United States. Their software collects a hybrid of clinical observations and provides clinical data-tracking and clinician feedback features that assists clinicians in improving health care. Patients were asked to finish the measurements before every psychotherapy session. The measurements include a Computer Adaptive Multidimensional Scale (CAMS) [28] and some commonly used psychotherapy scales for common disorders, such as the Generalized anxiety disorder (GAD) scale, the Patient Health Questionnaire (PHQ), and the Post Traumatic Stress Disorder Checklist (PCL).

The primary data used to demonstrate the analysis in the study consist of 402 participants[27](275 female and 127 male), who have received at least 3 treatments or visited clinics for 3 times since 2017. Their ages range from 18-69 ($M = 37.49$, $SD = 12.78$). For security, the data have been de-identified. Personally identifiable information, such as name and birth date, were removed by the data provider. The CAMS [28] measurement combines 70 questions on symptoms, functioning, and behaviors across 17 dimensions to produce a broad overview of an adult patient's mental health. Its 17 subscales measure Attachment, Avoidance, Connectedness, Hopelessness, Eating Problems, Emotional Distancing, Hurtful Rumination, Hypervigilance, Perfectionism, Pressure from Negative Affect, Psychosis, Resilience, Social Role, Relational Distress, Somatic Anxiety, Substance Use, and Suicide Risk. Each scale contains a screening question. On the first visit or treatment, a patient will be requested to finish all the questions. For subsequent visits, the screening question will be asked, but the following questions will only be asked if the screening question has a high score or if a running average of their previous scores is above a threshold. The adaptive feature of this measurement led to large number of missing values for many questions. Hence, only the 17 screening questions listed in Table 1 for which data were more complete, were selected for the analysis in current study. The data were split into train set(70%) and test set(30%) randomly.

3.1.1 Response(dependent) Variable – suicidal ideation

For each patient, the degree of agreement to the statement, "I think it would be better if I were dead.", was used as a response measure (dependent variable) of suicidal ideation and represented by Likert points, ranging from 1 to 7.

3.1.2 Independent Variables

Baseline demographic patient characteristics used for analysis included gender and age. The 7-point Likert scores to the other 15 questions at each time point were also included as independent variables in the study. Considering the response variable varies over time, time was included as an independent variable and measured by how many weeks the patient has been visiting.

3.2 Exploratory data analysis

To find a probability distribution that best fits the data, descriptive analyses were conducted with continuous variables as shown in Table 2.

Based on the descriptive analysis, the scores for suicidal ideation were rel-

Scale	Question
Attachment	I am someone who can form strong attachments to others.
Avoidance	I go out of my way to avoid certain places or experiences.
Connectedness	I have good friends who really know me.
Hopelessness	I feel things are too much for me to handle.
Eating	I spend far too much time and energy thinking about food and planning my meals.
Distancing	I can sometimes ”zone out” during intense or emotional moments.
Rumination	I can't stop worrying, even when I try.
Hypervigilance	I constantly think I need to be ready in case something bad happens.
Perfectionism	I need to feel in control at all times.
Negative Affect	I am feeling depressed.
Psychosis	I am very afraid that a secret organization is watching me.
Relational Distress	Other people don't seem to be able to understand me anyway, so I have given up.
Resilience	I'm good at letting others know what's important to me.
Social Role	Despite my difficulties, my community values me.
Somatic Anxiety	I feel restless and uneasy most of the time.
Substance Use	I am concerned that I am dependent on drinking and/or drugs.
Suicidal Ideation	I think it would be better if I were dead.

Table 1: Questions included in the analysis

atively low($M = 1.9$, $SD = 1.7$). The normality of the response variable was reviewed using the histogram and q-q plot shown in Figure 1.

Statistic	N	Mean	St. Dev.	Min	Max
Week Of Visit	3,416	12.8	12.4	0	51
Suicidal Ideation	2,959	1.9	1.7	1.0	7.0
Substance Use	2,833	1.6	1.5	1.0	7.0
Somatic Anxiety	2,905	4.1	1.9	1.0	7.0
Social Role	2,792	3.8	1.8	1.0	7.0
Resilience	2,887	4.3	1.8	1.0	7.0
Relational Distress	2,856	3.6	2.0	1.0	7.0
Psychosis	2,883	1.4	1.1	1.0	7.0
Negative Affect	2,920	4.3	2.0	1.0	7.0
Perfectionism	2,883	4.2	2.0	1.0	7.0
Stress	2,920	4.3	1.9	1.0	7.0
Hypervigilance	2,902	4.4	2.1	1.0	7.0
Rumination	2,930	4.5	1.9	1.0	7.0
Distancing	2,872	4.3	2.1	1.0	7.0
Eating	2,907	2.8	1.9	1.0	7.0
Connectedness	2,905	4.5	2.0	1.0	7.0
Avoidance	2,903	4.5	2.0	1.0	7.0
Attachment	2,888	4.8	1.8	1.0	7.0

Table 2: Descriptive Analysis of Continuous Variables

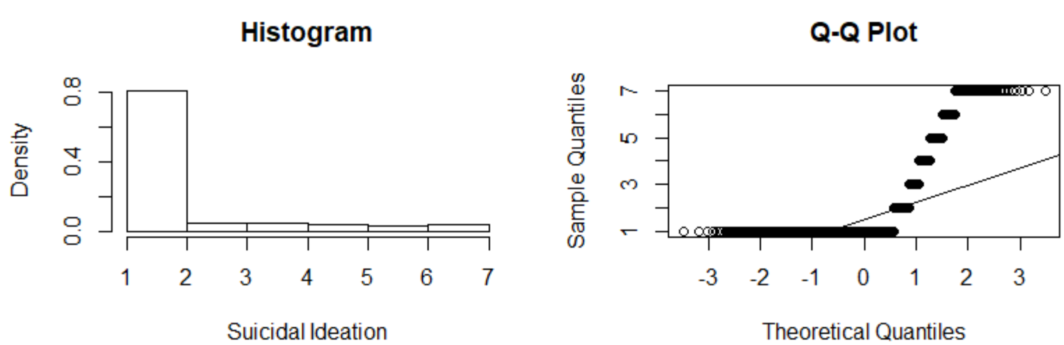


Figure 1: Assessing the normality of the response variable

To conduct GLMM, the distribution of the response variable needs to meet the assumption that variable needs to be normally distributed. When the histogram looks roughly bell-shaped and symmetric, or the Q-Q plots generally fall close to a diagonal line, we can conclude that the variable is normally distributed. However, based on Figure 1. The variable Suicidal Ideation failed to meet the requirements. A log-normal (or lognormal) distribution was used to test if the logarithm of the variable was normally distributed as shown in Figure 2. The y axis represents the observations and the x axis represents the quantiles modeled by the distribution. The solid line represents a perfect distribution fit and the dashed lines are the confidence intervals of the perfect distribution fit. In Figure 2, observations fell closer to the dashed lines. Therefore, the response variable was transformed using log-transformation for use in the GLMMs.

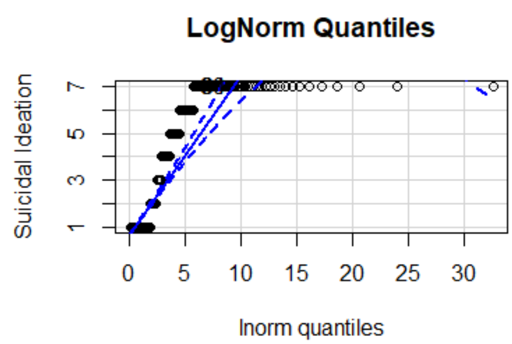


Figure 2: Fitting Log-Normal Distribution

To select the variables to be included in the model, a correlation matrix was calculated in Table 3.

	subjectAge	weekOfVisit	Suicidal Ideation	Substance Use	Somatic Anxiety	Social Role	Resilience	Relational Distress	Psychosis	Negative Affect
subjectAge										
weekOfVisit	0.03									
Suicidal Ideation	-0.1***	0.03								
Substance Use	-0.03	-0.13***	0.11***							
Somatic Anxiety	0.01	0.04	0.31***	-0.01						
Social Role	-0.08*	0.05	-0.18***	-0.04	-0.22***					
Resilience	0	0.09**	-0.15***	-0.09*	-0.07	0.38***				
Relational Distress	-0.06	0	0.45***	0	0.57***	-0.33***	-0.24***			
Psychosis	0.06	-0.01	0.15***	0.01	0.23***	0.01	0	0.22***		
Negative Affect	0.08*	0.04	0.44***	0.05	0.63***	-0.35***	-0.15***	0.52***	0.19***	
Perfectionism	0.01	0.04	0.25***	-0.02	0.41***	-0.04	0.04	0.3***	0.08*	0.33***
Hopelessness	-0.04	0	0.44***	0	0.55***	-0.27***	-0.13***	0.53***	0.09**	0.61***
Hypervigilance	0.02	0	0.25***	-0.07	0.54***	-0.1***	0.02	0.43***	0.19***	0.42***
Rumination	-0.02	-0.07	0.32***	0.02	0.67***	-0.25***	-0.09*	0.52***	0.15***	0.59***
Distancing	-0.05	0.02	0.26***	-0.03	0.53***	-0.19***	-0.08*	0.41***	0.2***	0.44***
Eating	0.04	-0.04	0.13***	0.09*	0.2***	-0.07	0.03	0.15***	0.06	0.24***
Connectedness	-0.01	0	-0.21***	-0.07	-0.22***	0.45***	0.48***	-0.36***	-0.01	-0.27***
Avoidance	0.11***	-0.06	0.22***	-0.01	0.48***	-0.14***	-0.07	0.4***	0.22***	0.4***
Attachment	-0.07	0.06	-0.15***	-0.09**	0.01	0.21***	0.43***	-0.14***	0.11***	0.03
Number of Visit	0.05	0.44***	0.16***	-0.19***	0.13***	0.01	0.13***	0.03	-0.01	0.2***
Suicidal Ideation_log	-0.08*	0	0.97***	0.11***	0.31***	-0.21***	-0.15***	0.44***	0.17***	0.45***
weekOfVsit_quadratic	0	0.95***	0.02	-0.12***	0.06	0.04	0.08*	0.01	-0.02	0.05

(a) Correlation Matrix

	Perfectionism	Hopelessness	Hypervigilance	Rumination	Distancing	Eating	Connectedness	Avoidance	Attachment	Number of Visit	Suicidal Ideation_log
Hopelessness	0.33***										
Hypervigilance	0.45***	0.42***									
Rumination	0.33***	0.59***	0.57***								
Distancing	0.32***	0.35***	0.41***	0.45***							
Eating	0.18***	0.29***	0.22***	0.18***	0.09*						
Connectedness	-0.02	-0.22***	-0.08*	-0.16***	-0.1***	-0.08*					
Avoidance	0.34***	0.38***	0.52***	0.49***	0.47***	0.13***	-0.13***				
Attachment	0.04	0.04	0.06	0.07	0.04	0.03	0.38***	0.06			
Number of Visit	0.17***	0.22***	0.13***	0.05	0.17***	0.05	0.03	0.08	0.13***		
Suicidal Ideation_log	0.24***	0.45***	0.26***	0.31***	0.24***	0.16***	-0.22***	0.2***	-0.14***	0.13***	
weekOfVsit_quadratic	0.03	0.02	0.01	-0.04	0.03	-0.03	-0.01	-0.05	0.08	0.38***	0

(b) Continued: Correlation Matrix

Note: * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

Table 3: Correlation Matrix

Based on the Table 3, variables, including subject age, Substance Use, Somatic Anxiety, Social Role, Resilience, Relation Distress, Psychosis, Negative Affect, Perfectionism, Hopelessness, Hypervigilance, Rumination, Distancing, Eating, Connectedness, Avoidance, Attachment, Number of Visit, were significantly correlated with Suicidal Ideation ($p \leq .001$). A time variable, "week of visit", was significantly correlated with Substance Use ($r = -.13, p \leq .001$), and Resilience ($r = .09, p \leq .01$), which were also significantly correlated with the quadratic form of time variable.

3.3 Generalized Linear Mixed Model

We utilized *R*[29] and *lme4* [30] to fit a generalized linear mixed model and perform the prediction with the test set. There are essentially two ways to fit a GLMM:

1. Starting with a small model and building up, or,

2. Starting with a big model and trimming down.

Considering the slated comparison with machine learning methods, we used the later strategy to proceed because it is more analogous to ML approach. All the variables that reflected significant correlations with Suicidal Ideation were added to the first baseline model as fixed effects, as well as the interaction forms of Time and Substance Use, Time and Resilience, Quadratic Time and Substance Use, and Quadratic Time and Resilience. The between-patient differences were included as random effects. The intercept for each patient was set as random.

As shown in Appendix A, variables were removed from the baseline model step by step when their p-values showed no significance ($p \leq .1$). First, the variable of participants' ages was removed with $\tilde{\chi}^2(1)$=.0027, $p = .958488 \gg .1$. The interaction form of quadratic time and Resilience was removed with $\tilde{\chi}^2(1)$=.0117, $p = .913915 \gg .1$. Next, the variables with lower significance were removed one by one following the order shown in Table 4, Eating ($\tilde{\chi}^2(1)$=.0145, $p = .904080 \gg .1$), Social Role ($\tilde{\chi}^2(1)$=.0431, $p = .835606 \gg .1$), Gender ($\tilde{\chi}^2(1)$=.0472, $p = .828014 \gg .1$), Resilience ($\tilde{\chi}^2(1)$=.1395, $p = .708772 \gg .1$), Interaction form of time and Resilience ($\tilde{\chi}^2(1)$=.8081, $p = .368677 \gg .1$), Avoidance ($\tilde{\chi}^2(1)$=1.1635, $p = .280749 \geq .1$), Rumination ($\tilde{\chi}^2(1)$=1.6697, $p = .196293 \geq .1$), Attachment ($\tilde{\chi}^2(1)$=1.0437, $p = .3069641 \geq .1$), Psychosis ($\tilde{\chi}^2(1)$=1.2339, $p = .266641 \geq .1$), Hypervigilance ($\tilde{\chi}^2(1)$=2.1446, $p = .142887 \geq .1$), Distancing ($\tilde{\chi}^2(1)$=3.4462, $p = .0633983 \geq .05$). Then, we have the final for GLMM:

$log(SuicidalIdeation) = weekofvisit * SubstanceUse + weekofvisit^2 * SubstanceUse + SomaticAnxiety + RelationalDistress + NegativeAffect + Perfectionism + Hopelessness + Connectedness + (1|subjectMirahId)$

The fixed effects included Somatic Anxiety, Relational Distress, Negative Affect, Perfectionism, Hopelessness, Connectedness, interaction form of time and Substance Use as well as an interaction form of quadratic time and Substance Use. As a random effect, we included the vector of random intercepts for subjects.

3.4 Machine Learning Methods: Random Forest(RF) and Gradient Boosting Decision Tree (GBDT)

The Random Forest(RF) model was selected for its robustness and performance in both accuracy and ease of implementation [31][32]. RF is a supervised Machine Learning algorithm for both regression and classification with the use of multiple decision trees and a technique called *Bagging*[33]. *Bagging*, in the RF method, involves training each decision tree on a different data sample where sampling is done with replacement. Then, multiple decision trees are combined to determine the final results as shown in Figure 3.

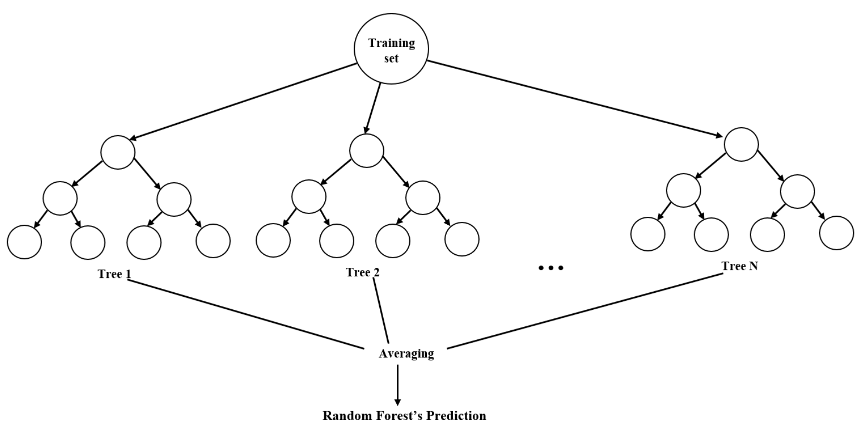


Figure 3: Random Forest Algorithm

Several recent studies have shown that classification models based on RF had relatively high accuracy when predicting suicidal ideation or suicide attempts[25][34]. We deployed the RF regression in our study to predict the scores for suicidal ideation.

Gradient boosting is one of powerful techniques for building predictive models[35]. Depending on the type of the problem, a loss function is used to optimize the model. The most basic structure of GBDT, as shown in Figure 4, is also decision trees. We will start with one decision tree, then trees are added one at a time stepwise. The parameters are modified and the results are combined at each step until we reach our goal of minimizing the loss to an acceptable level.

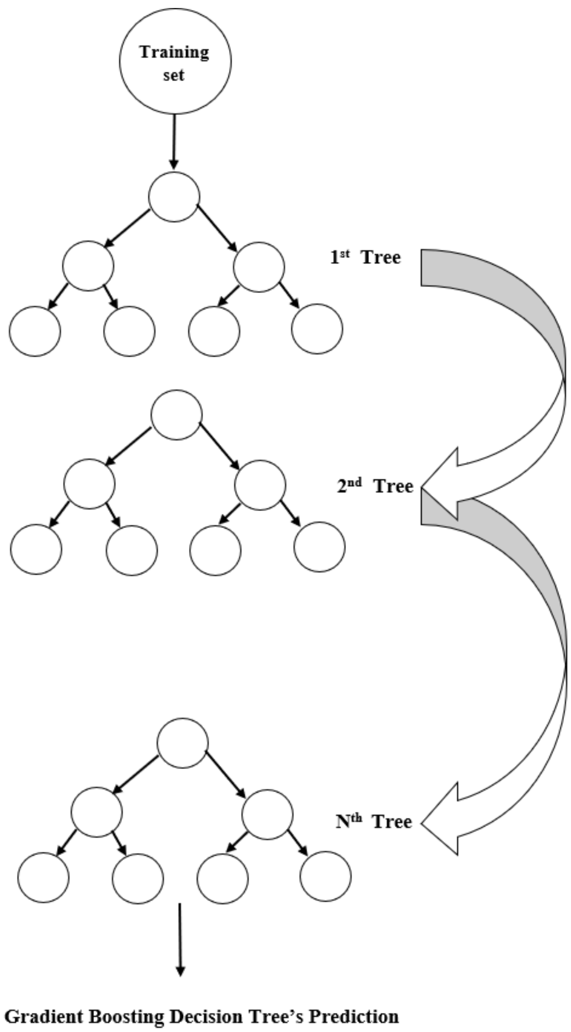


Figure 4: Gradient Boosting Decision Tree Algorithm

We used *Python*[36] and *Scikit-learn package* [37] to proceed RF regression and GBDT regression. All the fixed-effect variables were used to train the models with training set. We then used the grid search method from the *Scikit-learn* *package* [37] to determine the optimal values to be used for the hyperparameters of our models. Grid search is the process of performing hyper parameter tuning in order to determine the optimal values for a given model[38]. This method was also applied when training conditional RF and GBDT models - conditional models with variables selected from GLMM final version. Then, the best four models were deployed on test set to predict the score for suicidal ideation.

CHAPTER 4

Results

The fit of models and estimators of the predictors were compared among all the 14 Generalized Linear Mixed Models as shown in Table 4.

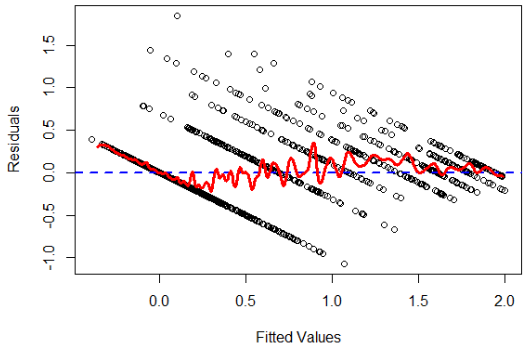
	#Df	AIC	BIC	LogLik	Df	Chisq	Pr(>Chisq)
Baseline Model	27	1076.1	1217.9	-634.34			
Dropping Subject Age	26	1076.1	1217.9	-629.22	-1	10.23	0.0014**
Dropping $time^2*$ Resilience	25	1074.1	1210.5	-619.59	-1	19.27	0.0000***
Dropping Eating	24	1072.1	1203.0	-615.48	-1	8.21	0.0042**
Dropping Social Role	23	1070.2	1195.6	-611.51	-1	7.93	0.0049**
Dropping Subject Gender	22	1068.2	1188.2	-609.75	-1	3.51	0.0609*
Dropping Resilience	22	1068.2	1188.2	-609.75	0	0.00	1.0000
Dropping $time*$ Resilience	20	1065.2	1174.3	-599.11	-2	21.29	0.0000***
Dropping Avoidance	19	1064.3	1168.0	-595.59	-1	7.05	0.0079**
Dropping Rumination	18	1064.0	1162.2	-592.46	-1	6.26	0.0124*
Dropping Attachment	17	1063.1	1155.8	-588.99	-1	6.94	0.0084**
Dropping Psychosis	16	1062.3	1149.6	-585.87	-1	6.24	0.0125*
Dropping Hypervigilance	15	1062.5	1144.3	-582.77	-1	6.18	0.0129*
Final Model(Dropping Distancing)	14	1063.9	1140.3	-580.31	-1	4.92	0.0265*

Notes: * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.
Final Model: $log(SuicidalIdeation) = weekofvisit * SubstanceUse + weekofvisit^2 * SubstanceUse + SomaticAnxiety + RelationalDistress + NegativeAffect + Perfectionism + Hopelessness + Connectedness + (1|subjectMirahId)$

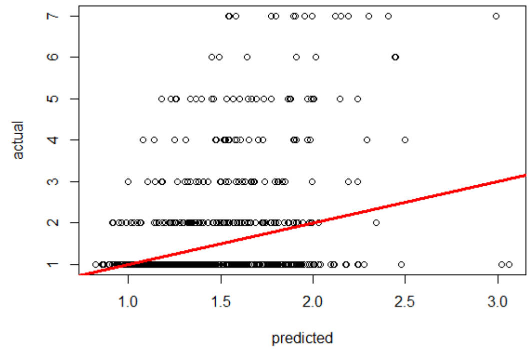
Table 4: Likelihood Ratio Test of 14 Models

The final Model showed significant improvement than the previous models, with (AIC = 1063.9, BIC = 1140.3).

We demonstrated the best fit models (Final GLMM, RF model, GBDT model, conditional RF model, and conditional GBDT model) with test data set to predict suicidal ideation. The performance of stet GLMM was displayed in residual plots in Figure 5.



(a) Residual Plot for GLMM



(b) Actual VS Predicted

Figure 5: Performance of GLMM

Figure 5(a) is a residual plot with predicted values on the x-axis and residuals on the y-axis. The distance from the line at 0 is how bad the prediction was for that value. The spots did not cluster towards the middle of the plot as shown in Figure 5(a). Figure 5(b) is a plot comparing the actual values of dependent variable with predicted values, in which there was not a strong correlation between the model's predictions and its actual values. Overall, the fit of the GLMM was poor. Then, the performance of all five models was evaluated using standard goodness of model fit (mean squared error(MSE), R^2, and mean absolute error(MAE)) as shown in Table 5.

Model	RSQUARE	MSE	MAE
GLMM	0.115	2.880	1.013
RF	0.357	1.915	0.980
GBDT	0.409	1.761	0.942
RF with selected variables*	0.260	2.202	0.975
GBDT with selected variables*	0.336	1.977	0.985

Notes: * variables in the final form of GLMM

Table 5: Model Comparison

Table 6 presents that overall, ML models performed better than GLMM when predicting suicidal ideation in test set. The GBDT models performed marginally better than RF models and the GBDT with all features included had the highest relative accuracy with $R^2 = .409$, MSE = 1.761 and MAE = .942.

RF and GBDT both consist of multiple decision trees. Each node in the decision tree is a condition of a single feature to separate the dataset into two sets. Within each set, the values for the dependent variable are similar. The measure based on which the optimal condition is chosen is called impurity, which is variance in regression trees[39]. Therefore, how much each feature decreases the weighted impurity will be computed when training a tree. Furthermore, when training a forest, the average impurity decrease from each feature can be calculated. And we used the term 'importance of the feature' to present the ranking of the features computed from this measure as shown in Table 6.

(a) Importances of the predictors in RF Model

feature	importance
Hopelessness	0.111844
Negative Affect	0.105749
subject Age	0.097951
Relational Distress	0.094780
Attachment	0.058146
Distancing	0.052875
time Of Visit	0.049620
Resilience	0.047253
week Of Visit	0.044720
Connectedness	0.040098
Perfectionism	0.037680
Eating	0.035467
Social Role	0.035078
Rumination	0.034871
Hypervigilance	0.031968
Somatic Anxiety	0.029871
Psychosis	0.028291
Avoidance	0.026586
Substance Use	0.022489
subject Gender	0.014665

(b) Importances of the predictors in GBDT Model

feature	importance
Hopelessness	0.215884
Negative Affect	0.182678
Relational Distress	0.170533
subject Age	0.075936
Attachment	0.058962
Resilience	0.033959
Somatic Anxiety	0.033269
Rumination	0.030351
Distancing	0.029263
Psychosis	0.027068
Eating	0.023441
Perfectionism	0.016457
time Of Visit	0.015496
Connectedness	0.015218
Social Role	0.013587
Hypervigilance	0.013424
Avoidance	0.013171
week Of Visit	0.012153
Substance Use	0.010437
subject Gender	0.008713

Table 6: Importances of the Predictors

As shown in Table 6, both the RF and GBDT models suggested that Hopelessness, Negative Affect, Subject Age, Relational Distress and Attachment were the top 5 most important features. However, as indicated in Appendix A., the GLMM model showed that Negative Affect, Substance Use, Hopelessness, Relational Distress and Connectedness were the most significant predictors. Surprisingly, Substance Use, which had second largest weight in GLMM, were the penultimate predictors in the RF and GBDT models. There were three overlapping factors, Hopelessness, Negative Affect and Relational Distress, which could be interpreted as strong risk factors when predicting suicidal ideation irregardless of model framework.

CHAPTER 5

Discussion

The present study was designed to assess and compare the performance of suicidal ideation prediction models built based on different algorithms. We applied GLMM, RF, and GBDT to partial EHR data in order to predict individuals with suicidal ideation in the population who are receiving psychotherapy and received psychotherapy. When predicting suicidal ideation in the test set ($n = 112$), the machine learning model, especially the GBDT model, showed a better performance than GLMM. Moreover, ML models helped to identify the predictors that are not significant in the GLMM, such as Attachment and Age. The common findings in these three models indicate that Hopelessness, Negative Affect and Relational Distress are important risk factors to predict individual's suicidal ideation. This finding has important implications for developing prediction models for suicide risk in the future.

The Machine Learning Method has better performance than Generalized Linear Mixed Models when developing prediction models based on a relatively large data set. It also bypasses the need for the traditional statistical procedure of Hypothesis creating-Model fitting-Hypothesis testing-p-value checking.

This study is subject to some methodological limitations. First, due to the adaptive feature of the measurement, most of the question data were missing by design. Thus we could only use 17 screening questions in our study, which might affect the training of the ML models. Second, we used limited amount of data for our model test in this study, which may introduce a higher testing error in the prediction models. Third, patients of our study were all involved in a certain kind of psychiatric treatment. We used counts of their visits as a dummy variable to indicate how many treatments they had received. This could lead to a biased result that was less interpretable. Fourth, the prediction models were based on the basic RF and GBDT algorithms without considering the variable selection and the mixed effect among variables. Advanced variable selection strategy and adding mixed effect might improve the prediction accuracy of the ML models. Lastly, we only deployed two decision-tree-based algorithms in the current study. Further analyses should be conducted to compare the performance of prediction models with additional machine learning algorithms, such as Support Vector Machines(SVM) and Neural Networks.

There are also some practical problems we need to concerned about. First, the sample used in the study was collected from one state in the US. In addition to assessing the prediction performance within this sample, we should consider generalizability to other care-provider locations or patient populations. Second, we need to be aware that prediction models cannot replace clinical judgment, but only provide an alert message to help with decision making and treatment modification. The models developed in this study were designed to address the question of who has the thought of killing himself/herself, but not when he/she will commit suicide. Decision-makers should be cautious when interpreting the predicted results with respect to patient safety.

In conclusion, this initial applied demonstration showed that Machine Learning(ML) models based on longitudinal EHR data could predict individuals with suicide ideation better than Generalized Linear Mixed Model(GLMM). Greater efforts are needed to apply additional ML algorithms to larger and more completed data sets to develop prediction models of more critical suicide risk, evaluated based on self-injurious behavior, suicide attempt, suicide plan and etc. In addition, systematic simulation studies comparing the two overall approaches for similar data set are needed in the future.

LIST OF REFERENCES

[1] W. H. Organization(WHO), "Suicide." Website, 2018, http://www.who.int/en/news-room/fact-sheets/detail/suicide.

[2] G. Arsenault-Lapierre, C. Kim, and G. Turecki, "Psychiatric diagnoses in 3275 suicides: a meta-analysis," *BMC psychiatry*, vol. 4, no. 1, p. 37, 2004.

[3] NIMH, "Serious mental illness (smi) among u.s. adults," Website, 2017, http://www.nimh.nih.gov/health/statistics/prevalence/serious-mental-illness-ami-among-adults.shtml.

[4] K. Hor and M. Taylor, "Suicide and schizophrenia: a systematic review of rates and risk factors," *Journal of psychopharmacology*, vol. 24, no. 4, pp. 81–90, 2010.

[5] C. Schoen, K. Davis, K. S. Collins, L. Greenberg, C. Des Roches, and M. Abrams, "The commonwealth fund survey of the health of adolescent girls," *New York: The Commonwealth Fund*, vol. 252, 1997.

[6] R. Miranda, A. Ortin, M. Scott, and D. Shaffer, "Characteristics of suicidal ideation that predict the transition to future suicide attempts in adolescents," *Journal of Child Psychology and Psychiatry*, vol. 55, no. 11, pp. 1288–1296, 2014.

[7] M. K. Nock, G. Borges, E. J. Bromet, J. Alonso, M. Angermeyer, A. Beautrais, R. Bruffaerts, W. T. Chiu, G. De Girolamo, S. Gluzman, *et al.*, "Cross-national prevalence and risk factors for suicidal ideation, plans and attempts," *The British Journal of Psychiatry*, vol. 192, no. 2, pp. 98–105, 2008.

[8] "What is an electronic health record (ehr)?" https://www.healthit.gov/faq/what-electronic-health-record-ehr.

[9] J. Zhao, A. Henriksson, L. Asker, and H. Boström, "redictive modeling of structured electronic health records for adverse drug event detection," *BMC Medical Informatics and Decision Making*, 2015.

[10] C. Safran, M. Bloomrosen, W. E. Hammond, S. Labkoff, S. Markel-Fox, P. C. Tang, and D. E. Detmer, "Toward a national framework for the secondary use of health data: an american medical informatics association white paper," *Journal of the American Medical Informatics Association*, vol. 14, no. 1, pp. 1–9, 2007.

[11] G. E. Simon, E. Johnson, J. M. Lawrence, R. C. Rossom, B. Ahmedani, F. L. Lynch, A. Beck, B. Waitzfelder, R. Ziebell, R. B. Penfold, and S. M. Shortreed, "Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records," *American Journal of Psychiatry*, vol. 175, no. 10, pp. 951–960, 2018.

[12] W. W. Stroup, *Generalized linear mixed models: modern concepts, methods and applications.* CRC press, 2016.

[13] B. M. Bolker, M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White, "Generalized linear mixed models: a practical guide for ecology and evolution," *Trends in ecology & evolution*, vol. 24, no. 3, pp. 127–135, 2009.

[14] T. A. Snijders, R. J. Bosker, *et al.*, "An introduction to basic and advanced multilevel modeling," *Sage, London. WONG, GY, y MASON, WM (1985): The Hierarchical Logistic Regression. Model for Multilevel Analysis, Journal of the American Statistical Association*, vol. 80, no. 5, pp. 13–524, 1999.

[15] L. Huang, L. Tang, B. Zhang, Z. Zhang, and H. Zhang, "Comparison of different computational implementations on fitting generalized linear mixed-effects models for repeated count measures," *Journal of Statistical Computation and Simulation*, vol. 86, no. 12, pp. 2392–2404, 2016.

[16] A. Gelman and J. Hill, "Data analysis using regression and hierarchical/multilevel models," *New York, NY: Cambridge*, 2007.

[17] A. Zuur, E. N. Ieno, N. Walker, A. A. Saveliev, and G. M. Smith, *Mixed effects models and extensions in ecology with R.* Springer Science & Business Media, 2009.

[18] C. Ngufor, H. Van Houten, B. S. Caffo, N. D. Shah, and R. G. McCoy, "Mixed effect machine learning: A framework for predicting longitudinal change in hemoglobin a1c," *Journal of biomedical informatics*, vol. 89, pp. 56–67, 2019.

[19] J. Cohen, P. Cohen, S. West, and L. Aiken, "Applied multiple regression/correlation analysis for the behavioral sciences. lawrence erlbaum associates," *Mahwah, NJ*, 2003.

[20] J. R. Koza, D. E. Goldberg, D. B. Fogel, and R. L. Riolo, *Genetic Programming 1996: proceedings of the first annual conference.* Mit Press, 1996.

[21] T. Yarkoni and J. Westfall, "Choosing prediction over explanation in psychology: Lessons from machine learning," *Perspectives on Psychological Science*, vol. 12, no. 6, pp. 1100–1122, 2017.

[22] C. Strobl, J. Malley, and G. Tutz, "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests." *Psychological methods*, vol. 14, no. 4, p. 323, 2009.

[23] A. J. Rosellini, J. Monahan, A. E. Street, E. D. Hill, M. Petukhova, B. Y. Reis, N. A. Sampson, D. M. Benedek, P. Bliese, M. B. Stein, *et al.*, "Using administrative data to identify us army soldiers at high-risk of perpetrating minor violent crimes," *Journal of psychiatric research*, vol. 84, pp. 128–136, 2017.

[24] T. Ding, W. K. Bickel, and S. Pan, "Multi-view unsupervised user feature embedding for social media-based substance use prediction," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 2275–2284.

[25] C. G. Walsh, J. D. Ribeiro, and J. C. Franklin, "Predicting risk of suicide attempts over time through machine learning," *Clinical Psychological Science*, vol. 5, no. 3, pp. 457–469, 2017.

[26] Y. Barak-Corren, V. M. Castro, S. Javitt, A. G. Hoffnagle, Y. Dai, R. H. Perlis, M. K. Nock, J. W. Smoller, and B. Y. Reis, "Predicting suicidal behavior from longitudinal electronic health records," *American journal of psychiatry*, vol. 174, no. 2, pp. 154–162, 2016.

[27] "Mirah corporation," https://www.mirah.com/.

[28] C. Moltu, M. Veseth, J. Stefansen, J. C. Nøtnes, Å. Skjølberg, P.-E. Binder, L. G. Castonguay, and S. S. Nordberg, "This is what i need a clinical feedback system to do for me: a qualitative inquiry into therapists' and patients' perspectives," *Psychotherapy research*, vol. 28, no. 2, pp. 250–263, 2018.

[29] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org/

[30] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[31] R. C. Kessler, H. M. van Loo, K. J. Wardenaar, R. M. Bossarte, L. A. Brenner, T. Cai, D. D. Ebert, I. Hwang, J. Li, P. de Jonge, *et al.*, "Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports," *Molecular psychiatry*, vol. 21, no. 10, p. 1366, 2016.

[32] J. Futoma, J. Morris, and J. Lucas, "A comparison of models for predicting early hospital readmissions," *Journal of biomedical informatics*, vol. 56, pp. 229–238, 2015.

[33] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[34] S. Ryu, H. Lee, D.-K. Lee, and K. Park, "Use of a machine learning algorithm to predict individuals with suicide ideation in the general population," *Psychiatry investigation*, vol. 15, no. 11, p. 1030, 2018.

[35] J. H. Friedman, "Stochastic gradient boosting," *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[36] G. van Rossum, "Python tutorial," Centrum voor Wiskunde en Informatica (CWI), Amsterdam, Tech. Rep. CS-R9526, May 1995.

[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[38] Krishni, "An introduction to grid search," Website, 2019, https://medium.com/datadriveninvestor/an-introduction-to-grid-search-ff57adcc0998.

[39] A. Saabas, "Selecting good features," Website, 2014, https://blog.datadive.net/selecting-good-features-part-iii-random-forests/.

	Overall Rating						
	Baseline Model	Dropping subject Age	Dropping $time^2 * Resilience$	Dropping Eating	Dropping Social Role	Dropping subject Gender	Dropping Resilience
subjectAge	0.0001						
	(0.002)						
subjectGender	−0.015	-0.014	-0.014	-0.015	-0.015		
	(0.069)	(0.067)	(0.067)	(0.067)	(0.067)		
weekOfVisit	−0.010	−0.010	−0.009***	−0.009***	−0.009***	−0.009***	−0.009***
	(0.006)	(0.006)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Substance Use	0.026**	0.026**	0.026**	0.026**	0.026**	0.026**	0.026**
	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
$WeekOfVisit^2$	0.0002	0.0002	0.0002***	0.0002***	0.0002***	0.0002***	0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Somatic Anxiety	0.011	0.011	0.011	0.011	0.011	0.011	0.011
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Social Role	−0.001	−0.002	−0.002	−0.002			
	(0.007)	(0.007)	(0.007)	(0.007)			
Resilience	0.001	0.001	0.002	0.002	0.002	0.002	0.002
	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Relational Distress	0.016**	0.016**	0.016**	0.016**	0.016**	0.016**	0.016**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Psychosis	0.011	0.011	0.011	0.011	0.011	0.011	0.011
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Negative Affect	0.038***	0.039***	0.039***	0.038***	0.039***	0.039***	0.039***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Perfectionism	0.011*	0.011*	0.011*	0.011*	0.011*	0.011*	0.011*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Hopelessness	0.021***	0.021***	0.021***	0.021***	0.021***	0.022***	0.022***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Hypervigilance	0.008	0.008	0.008	0.008	0.008	0.008	0.008
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Rumination	0.010	0.010	0.010	0.010	0.010	0.010	0.010
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Distancing	−0.012*	−0.012*	−0.012*	−0.012*	−0.012*	−0.012*	−0.012*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Eating	0.001	0.001	0.001				
	(0.007)	(0.007)	(0.007)				
Connectedness	−0.015**	−0.015**	−0.015**	−0.015**	−0.016**	−0.016**	−0.016**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Avoidance	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Attachment	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006
	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
$weekOfVisit * SubstanceUse$	0.004***	0.004***	0.004***	0.004***	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$SubstanceUse * weekOfVisit^2$	−0.0001**	−0.0001**	−0.0001**	−0.0001**	−0.0001**	−0.0001**	−0.0001**
	(0.00005)	(0.00005)	(0.00005)	(0.00005)	(0.00005)	(0.00005)	(0.00005)
$weekOfVisit * Resilience$	−0.0002	−0.0002	−0.0003	−0.0003	−0.0003	−0.0003	−0.0003
	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
$Resilience * weekOfVisit^2$	−0.00000	−0.00000					
	(0.00003)	(0.00003)					
Constant	0.012	0.016	0.014	0.015	0.011	0.006	0.006
	(0.126)	(0.085)	(0.081)	(0.081)	(0.079)	(0.076)	(0.076)
Observations	1,730	1,730	1,730	1,730	1,730	1,730	1,730
Akaike Inf. Crit.	1,322.674	1,310.442	1,289.171	1,278.958	1,269.023	1,263.510	1,263.510
Bayesian Inf. Crit.	1,469.982	1,452.295	1,425.568	1,409.899	1,394.508	1,383.539	1,383.539
Note:	*p<0.1; **p<0.05; ***p<0.01						

Table A1: GLMM Model Fit

	Overall Rating						
	Dropping $time * resilience$	Dropping Avoidance	Dropping Rumination	Dropping Attachment	Dropping Psychosis	Dropping Hypervigilance	Dropping Distancing
subjectAge							
subjectGender							
weekOfVisit	−0.010***	−0.010***	−0.010***	−0.010***	−0.010***	−0.010***	−0.010***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Substance Use	0.025**	0.026**	0.026**	0.026**	0.027**	0.028**	0.028**
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
$WeekOfVisit^2$	0.0002***	0.0002***	0.0002***	0.0002***	0.0002***	0.0002***	0.0002**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Somatic Anxiety	0.011	0.011	0.013*	0.013*	0.014**	0.015**	0.013*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Social Role	−0.001	−0.002	−0.002	−0.002			
Resilience							
Relational Distress	0.016**	0.016**	0.017**	0.017**	0.018***	0.019***	0.018***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Psychosis	0.010	0.010	0.010	0.011			
	(0.010)	(0.010)	(0.010)	(0.010)			
Negative Affect	0.039***	0.039***	0.040***	0.040***	0.041***	0.041***	0.040***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Perfectionism	0.011*	0.010	0.011*	0.011*	0.011*	0.013**	0.012*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Hopelessness	0.022***	0.021***	0.022***	0.021***	0.021***	0.022***	0.022***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Hypervigilance	0.008	0.007	0.009	0.009	0.009		
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)		
Rumination	0.010	0.010					
	(0.008)	(0.008)					
Distancing	−0.012*	−0.013**	−0.012**	−0.013**	−0.012**	−0.011*	
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	
Eating							
Connectedness	−0.016**	−0.016**	−0.015**	−0.017**	−0.016**	−0.016**	−0.016**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Avoidance	−0.007						
	(0.007)						
Attachment	−0.007	−0.008	−0.008				
	(0.007)	(0.007)	(0.007)				
$weekOfVisit * SubstanceUse$	0.004***	0.004***	0.004***	0.004***	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$SubstanceUse * weekOfVisit^2$	−0.0001**	−0.0001**	−0.0001*	−0.0001*	−0.0001*	−0.0001*	−0.0001*
	(0.00005)	(0.00005)	(0.00005)	(0.00005)	(0.00005)	(0.00005)	(0.00005)
$weekOfVisit * canCommunicateWhatIsImportant$							
$canCommunicateWhatIsImportant * weekOfVisit^2$							
Constant	0.019	0.007	0.012	−0.013	−0.012	−0.003	−0.027
	(0.071)	(0.070)	(0.070)	(0.066)	(0.066)	(0.065)	(0.064)
Observations	1,730	1,730	1,730	1,730	1,730	1,730	1,730
Akaike Inf. Crit.	1,238.224	1,229.173	1,220.915	1,211.974	1,203.734	1,195.550	1,188.629
Bayesian Inf. Crit.	1,347.342	1,332.835	1,319.121	1,304.724	1,291.028	1,277.388	1,265.011
Note:	*p<0.1; **p<0.05; ***p<0.01						

Table A2: Continued: GLMM Model Fit

BIBLIOGRAPHY

"Mirah corporation," https://www.mirah.com/.

"What is an electronic health record (ehr)?" https://www.healthit.gov/faq/what-electronic-health-record-ehr.

Arsenault-Lapierre, G., Kim, C., and Turecki, G., "Psychiatric diagnoses in 3275 suicides: a meta-analysis," *BMC psychiatry*, vol. 4, no. 1, p. 37, 2004.

Barak-Corren, Y., Castro, V. M., Javitt, S., Hoffnagle, A. G., Dai, Y., Perlis, R. H., Nock, M. K., Smoller, J. W., and Reis, B. Y., "Predicting suicidal behavior from longitudinal electronic health records," *American journal of psychiatry*, vol. 174, no. 2, pp. 154–162, 2016.

Bates, D., Mächler, M., Bolker, B., and Walker, S., "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., and White, J.-S. S., "Generalized linear mixed models: a practical guide for ecology and evolution," *Trends in ecology & evolution*, vol. 24, no. 3, pp. 127–135, 2009.

Breiman, L., "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

Cohen, J., Cohen, P., West, S., and Aiken, L., "Applied multiple regression/correlation analysis for the behavioral sciences. lawrence erlbaum associates," *Mahwah, NJ*, 2003.

Ding, T., Bickel, W. K., and Pan, S., "Multi-view unsupervised user feature embedding for social media-based substance use prediction," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 2275–2284.

Friedman, J. H., "Stochastic gradient boosting," *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367–378, 2002.

Futoma, J., Morris, J., and Lucas, J., "A comparison of models for predicting early hospital readmissions," *Journal of biomedical informatics*, vol. 56, pp. 229–238, 2015.

Gelman, A. and Hill, J., "Data analysis using regression and hierarchical/multilevel models," *New York, NY: Cambridge*, 2007.

Hor, K. and Taylor, M., "Suicide and schizophrenia: a systematic review of rates and risk factors," *Journal of psychopharmacology*, vol. 24, no. 4, pp. 81–90, 2010.

Huang, L., Tang, L., Zhang, B., Zhang, Z., and Zhang, H., "Comparison of different computational implementations on fitting generalized linear mixed-effects models for repeated count measures," *Journal of Statistical Computation and Simulation*, vol. 86, no. 12, pp. 2392–2404, 2016.

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Cai, T., Ebert, D. D., Hwang, I., Li, J., de Jonge, P., *et al.*, "Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports," *Molecular psychiatry*, vol. 21, no. 10, p. 1366, 2016.

Koza, J. R., Goldberg, D. E., Fogel, D. B., and Riolo, R. L., *Genetic Programming 1996: proceedings of the first annual conference*. Mit Press, 1996.

Krishni, "An introduction to grid search," Website, 2019, https://medium.com/datadriveninvestor/an-introduction-to-grid-search-ff57adcc0998.

Miranda, R., Ortin, A., Scott, M., and Shaffer, D., "Characteristics of suicidal ideation that predict the transition to future suicide attempts in adolescents," *Journal of Child Psychology and Psychiatry*, vol. 55, no. 11, pp. 1288–1296, 2014.

Moltu, C., Veseth, M., Stefansen, J., Nøtnes, J. C., Skjølberg, Å., Binder, P.-E., Castonguay, L. G., and Nordberg, S. S., "This is what i need a clinical feedback system to do for me: a qualitative inquiry into therapists' and patients' perspectives," *Psychotherapy research*, vol. 28, no. 2, pp. 250–263, 2018.

Ngufor, C., Van Houten, H., Caffo, B. S., Shah, N. D., and McCoy, R. G., "Mixed effect machine learning: A framework for predicting longitudinal change in hemoglobin a1c," *Journal of biomedical informatics*, vol. 89, pp. 56–67, 2019.

NIMH, "Serious mental illness (smi) among u.s. adults," Website, 2017, http://www.nimh.nih.gov/health/statistics/prevalence/serious-mental-illness-ami-among-adults.shtml.

Nock, M. K., Borges, G., Bromet, E. J., Alonso, J., Angermeyer, M., Beautrais, A., Bruffaerts, R., Chiu, W. T., De Girolamo, G., Gluzman, S., *et al.*, "Cross-national prevalence and risk factors for suicidal ideation, plans and attempts," *The British Journal of Psychiatry*, vol. 192, no. 2, pp. 98–105, 2008.

Organization(WHO), W. H., "Suicide." Website, 2018, http://www.who.int/en/news-room/fact-sheets/detail/suicide.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org/

Rosellini, A. J., Monahan, J., Street, A. E., Hill, E. D., Petukhova, M., Reis, B. Y., Sampson, N. A., Benedek, D. M., Bliese, P., Stein, M. B., *et al.*, "Using administrative data to identify us army soldiers at high-risk of perpetrating minor violent crimes," *Journal of psychiatric research*, vol. 84, pp. 128–136, 2017.

Ryu, S., Lee, H., Lee, D.-K., and Park, K., "Use of a machine learning algorithm to predict individuals with suicide ideation in the general population," *Psychiatry investigation*, vol. 15, no. 11, p. 1030, 2018.

Saabas, A., "Selecting good features," Website, 2014, https://blog.datadive.net/selecting-good-features-part-iii-random-forests/.

Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C., and Detmer, D. E., "Toward a national framework for the secondary use of health data: an american medical informatics association white paper," *Journal of the American Medical Informatics Association*, vol. 14, no. 1, pp. 1–9, 2007.

Schoen, C., Davis, K., Collins, K. S., Greenberg, L., Des Roches, C., and Abrams, M., "The commonwealth fund survey of the health of adolescent girls," *New York: The Commonwealth Fund*, vol. 252, 1997.

Simon, G. E., Johnson, E., Lawrence, J. M., Rossom, R. C., Ahmedani, B., Lynch, F. L., Beck, A., Waitzfelder, B., Ziebell, R., Penfold, R. B., and Shortreed, S. M., "Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records," *American Journal of Psychiatry*, vol. 175, no. 10, pp. 951–960, 2018.

Snijders, T. A., Bosker, R. J., *et al.*, "An introduction to basic and advanced multilevel modeling," *Sage, London. WONG, GY, y MASON, WM (1985): The Hierarchical Logistic Regression. Model for Multilevel Analysis, Journal of the American Statistical Association*, vol. 80, no. 5, pp. 13–524, 1999.

Strobl, C., Malley, J., and Tutz, G., "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests." *Psychological methods*, vol. 14, no. 4, p. 323, 2009.

Stroup, W. W., *Generalized linear mixed models: modern concepts, methods and applications.* CRC press, 2016.

van Rossum, G., "Python tutorial," Centrum voor Wiskunde en Informatica (CWI), Amsterdam, Tech. Rep. CS-R9526, May 1995.

Walsh, C. G., Ribeiro, J. D., and Franklin, J. C., "Predicting risk of suicide attempts over time through machine learning," *Clinical Psychological Science*, vol. 5, no. 3, pp. 457–469, 2017.

Yarkoni, T. and Westfall, J., "Choosing prediction over explanation in psychology: Lessons from machine learning," *Perspectives on Psychological Science*, vol. 12, no. 6, pp. 1100–1122, 2017.

Zhao, J., Henriksson, A., Asker, L., and Boström, H., "redictive modeling of structured electronic health records for adverse drug event detection," *BMC Medical Informatics and Decision Making*, 2015.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., and Smith, G. M., *Mixed effects models and extensions in ecology with R.* Springer Science & Business Media, 2009.