iSense: AI-based Early Detection Tool to Identify Linguistic Bio-Markers of Mood Disorders and Recognize At-Risk Teens

INTRODUCTION

- Every 100 minutes, another teen takes their life, often with no perceivable warning. Suicide is the third-leading cause of death in young people age 15 to 19, and the key to reducing the prevalence lies in detecting its top risk factor: mental health disorders, particularly depression and other mood disorders

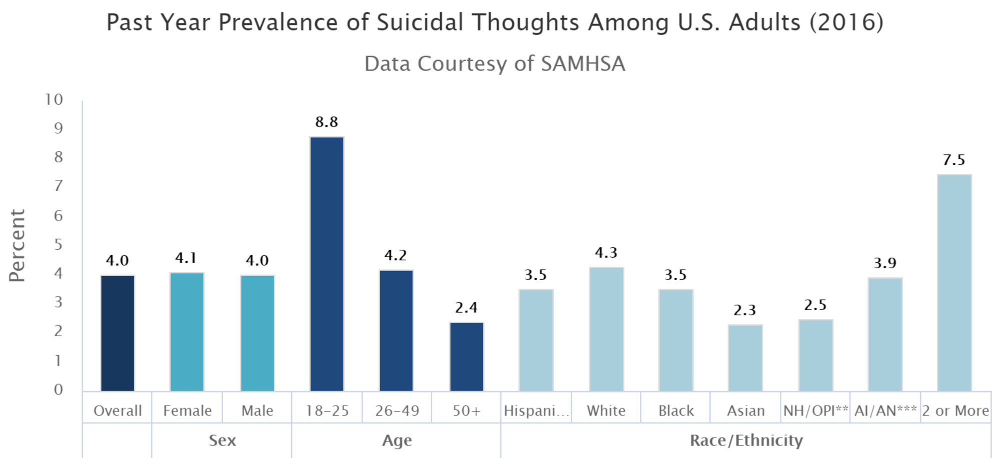


Figure 1

Source: National Institue of Mental Health

- Mental health disorders affect over 450 million people worldwide, more than cancer, diabetes, or heart disease
- As shown in **Figure 1**, 8.8% of 18-25 year olds (~6 million people) are currently struggling with a mental health disorder
- 49% of mental health patients begin manifesting symptoms by age 18, but despite the early onset, between 35 to 50 percent of individuals in high income countries are not diagnosed
- The prolonged delay between the initial appearance of symptoms and the time of diagnosis (if ever diagnosed) is alarming and allows for the condition to become increasingly severe, leading to possible suicidal behavior
- A cost-effective, efficient, and accurate solution must be developed to combat the growing mental health epidemic and increase in suicide cases

MOOD DISORDERS

- A category of mental illness in which the underlying problem primarily affects a person's persistent emotional state
- Can be classified into two broad groups: unipolar (depressive disorders) and bipolar disorders (manic depression), as seen in **Figure 2**
- This study focuses on the presence of depression and its top risk factors as described by the DSM-5 Criteria, whether it be part of a unipolar depressive disorder or the depressed period of a bipolar disorder

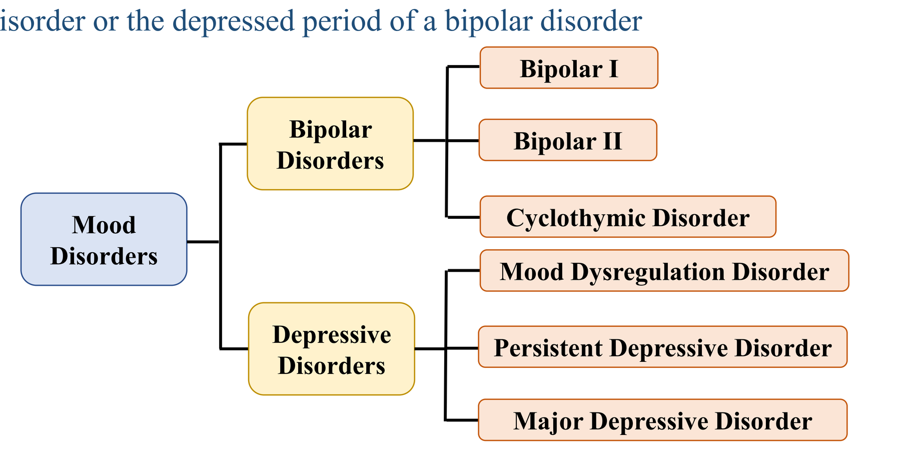


Figure 2

CURRENT SOLUTIONS

- In 2016, school guidance counselors in most states implemented mandated suicide awareness programs in high schools and assured students that their door is open. However, the likelihood that a teen who is seriously considering suicide would speak to a counselor is small
- Other initiatives recently taken in schools are suicide risk screenings, at-risk student referrals, and crisis emergency responses
- Misdiagnosis rates reach ~97.2% for bipolar disorder and ~65.9% for major depressive disorder

PROFESSIONALS' PERSPECTIVE

Mental Health Professional's Perspective

CEO of Metropolitan Counseling Services, mentioned the following:

- Health insurance does not typically cover mental healthcare, and the average therapy session costs $150-175 per hour (most patients attend once a week)
- The longer a patient waits, whether it be due to social stigma or the cost, the more therapy they will need to fully recover. The average teen waits 10 years before receiving treatment
- This app can increase early detection of mood disorders, and the sooner that the disorder is detected, the less therapy the individual will ultimately need. This will also reduce cost for families significantly

School Counselor's Perspective

- Currently, the process for identifying students who may have a mood disorder includes teachers observing deviations from the norm, parents calling in, and students talking to counselors themselves
- The school partners with a mental health organization, and a counselor visits every week to talk to kids who may not be able to afford private mental health care

GENERAL PROJECT INFORMATION

Research Question: Can a mathematical model identify and quantify the relationship between a healthy individual's language use in comparison to that of an individual with a mood disorder?

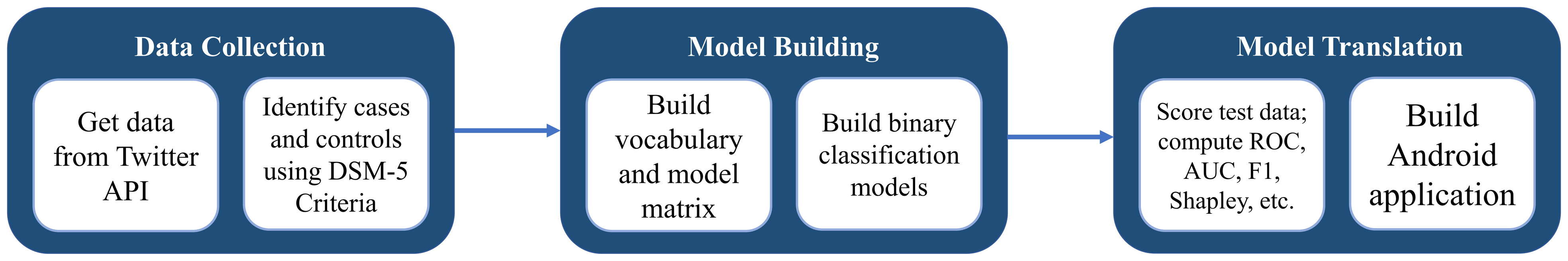
Hypothesis: If a healthy individual's language use is compared to that of an individual with a mood disorder, then there exists a difference in terms of linguistic biomarkers which can be identified by a mathematical model.

Independent Variable: Linguistic Biomarkers

- Specific n-grams (words or groups of words)
- Structure (indefinite pronouns, self-referent word usage, etc.)
- Certain topics (reference to drugs, display of outward insecurity, etc.)
- Other linguistic implications determined by the model

Dependent Variable: Mental state of the individual

METHODS



Data Collection

- Tweepy is an open-source library that enables Python to communicate with the Twitter API
- A script which utilizes these capabilities was run for seven days, and over 1.5 million tweets and their corresponding twitter handles were collected in real time

Depressive Disorders DSM-5 Diagnostic Criteria

Criteria	Presence in Tweet/Account
Depressed mood most of the day, nearly every day	User tweets a message indicative of a mood disorder at least three times a day for a two week period
Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day	User does not tweet about their occupation, education, activities, etc. that they previously tweeted about, and these tweets are absent for a two week period
Significant weight loss or gain when not dieting, or decrease in appetite nearly every day	Cannot determine from tweet
A slowing down of thought and reduction of physical movement	Cannot determine from tweet
Fatigue or loss of energy nearly every day	User tweets messages expressing fatigue at least ten times over a two week period
Feelings of worthlessness or excessive/inappropriate guilt nearly every day	User tweets messages apologizing excessively at least ten times over a two week period
Diminished ability to think or concentrate, indecisiveness, nearly every day	Cannot determine from tweet
Recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or a suicide attempt/specific plan for commiting suicide	User tweets a message discussing or idealizing suicide at least five times over a two week period

- Over 800 case and control individuals were identified based on the DSM-5 Diagnostic Criteria
- After a second round of filtering, the ground truth dataset consisted of 73,944 tweets

Model Building

- Tweets were divided into n-grams (up to 5-grams were tried), which were stored in a vocabulary
- Only n-grams that are present in a sufficient fraction (0.1%) of tweets were included in the final vocabulary
- The dataset was split into train (80%) and test (20%) groups, train data was split into 4 folds, and train and test sparse matrices were created
- Three models trained on the sparse matrix: Generalized Linear Model, Gradient Boosting Machine, and Multilayer Perceptron

Generalized Linear Model	Gradient Boosting Machine	Multilayer Perceptron
• Hyperplane-based approach from H_2O library • Regularization path created to select features for model • 4-fold cross validation to avoid overfitting	• Tree-based approach • Differs from other tree-based approaches because tree grows leaf-wise, not level-wise • Several combinations of parameters tried to determine optimal values	• Deep learning-based approach • A class of Feed-Forward Artificial Neural Networks containing an input layer, hidden layers, and an output layer

Model Translation

- Accuracy, sensitivity (true positive rate), specificity (true negative rate), and F1 score were analyzed as the threshold between case and control changed
- Optimal threshold was determined based on these values
- Model implemented in the mobile application

RESULTS – GENERALIZED LINEAR MODEL

Distribution of Scores

- **Figure 3** shows the number of documents in each score bin by percentage of observations per class
- In model training/testing, the mean score of the cases was 0.51 and the standard deviation was 0.16, while the mean score of the controls was 0.36 and the standard deviation was 0.11

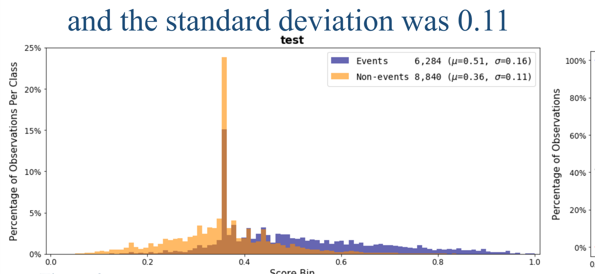


Figure 3

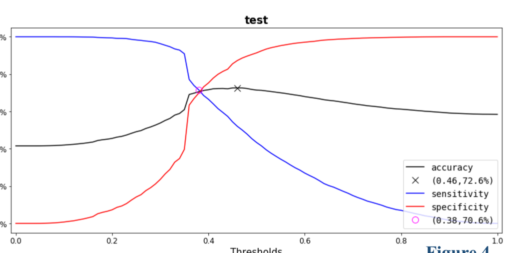


Figure 4

Determining Optimal Threshold

- **Figure 4** shows accuracy, sensitivity, and specificity as the score threshold between case and control changes
- In testing, 0.46 was the the score threshold with maximum accuracy, and 0.38 was the score threshold where all three measures are optimal

Model Accuracy

- The Receiver Operating Characteristic curve (**Figure 5**) plots the False Positive Rate (1 – specificity) against the True Positive Rate (sensitivity)
- Model training had an AUC of 0.7892 and model testing had an AUC of 0.7860, indicating that the model did not overfit to the training data

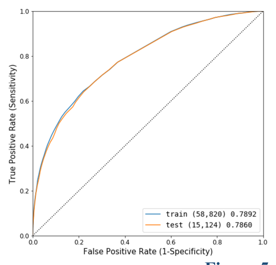


Figure 5

Important Features

- **Figure 6** shows the linguistic biomarkers that were determined to be the most predictive of mood disorders by the GLM model
- The intercept value was -0.672
- Negative coefficients were also returned, indicating a protective effect

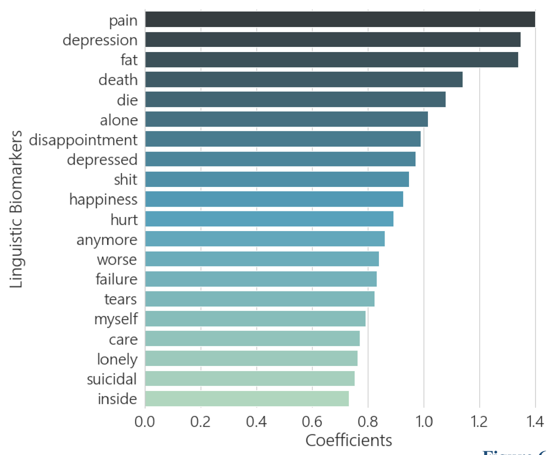


Figure 6

RESULTS – GRADIENT BOOSTING MACHINE

Model Accuracy/Important Features

- **Figure 7** shows that model training had an AUC of 0.8213 and model testing had an AUC of 0.8092, indicating that the model did not overfit to the training data
- **Figure 8** shows that 0.7665 is the average precision across all sensitivities in test
- **Figure 9** shows the highest shapley values for linguistic biomarkers in the top 200 tweets with the highest score

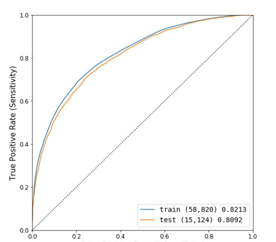


Figure 7

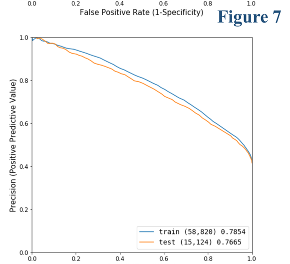


Figure 8

Shapley values, a method originating from game theory, help quantify the contribution of each feature in a prediction. The Shapley value is the average marginal contribution of a feature value across all possible feature combinations or coalitions.

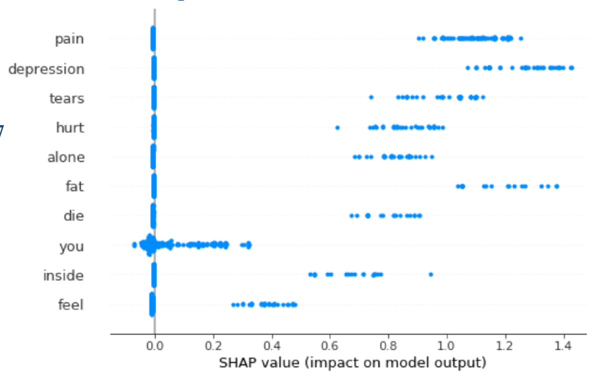


Figure 9

Shapley Values

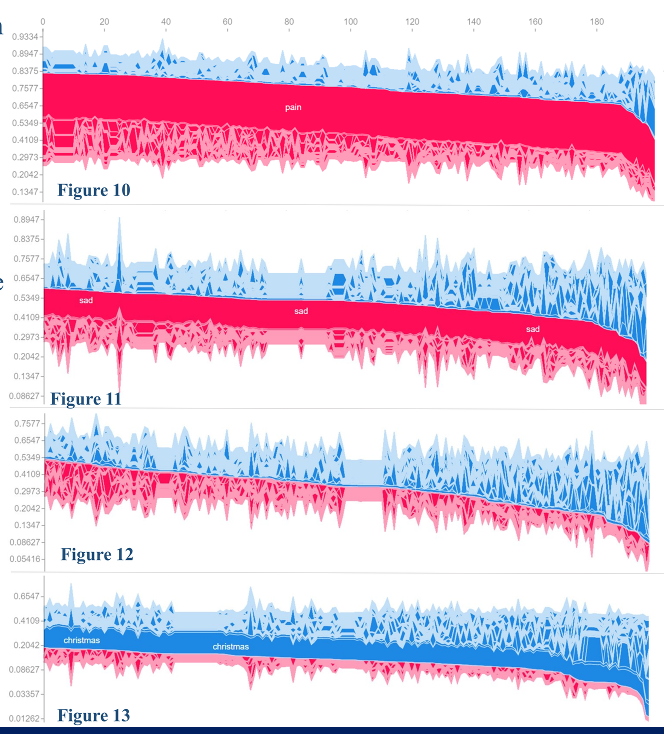


Figure 10

Figure 11

Figure 12

Figure 13

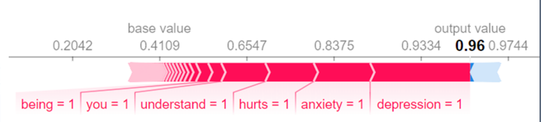


Figure 14

- **Figures 10-13** show that while tweets with "pain" and "sad" are almost always predicted as cases, tweets with "family" can be cases or controls, depending on other words used, and tweets with "Christmas" are almost always controls
- **Figures 14 and 15** show shapley values of words in two example tweets (impact of the word on the score is indicated by the length of the word's bar)

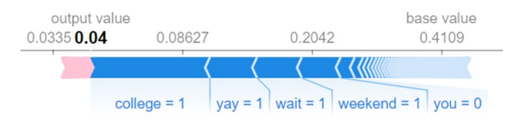


Figure 15

RESULTS – MULTILAYER PERCEPTRON

Network Architecture

- **Figures 16 and 17** show the architecture of the feed-forward deep neural network: an input layer, an output layer, and two hidden layers
- 991 features (linguistic biomarkers) are in the input of the first layer
- 52,201 total parameters are computed between all layers, and one score is returned
- The Rectified Linear Unit (ReLU) activation function is used for the hidden layers, and the Sigmoid activation function is used for the output layer

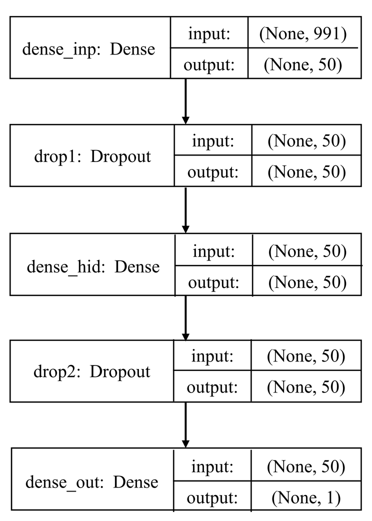


Figure 16

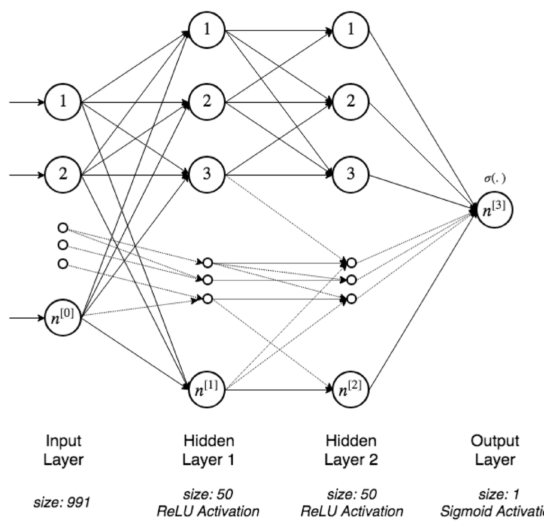
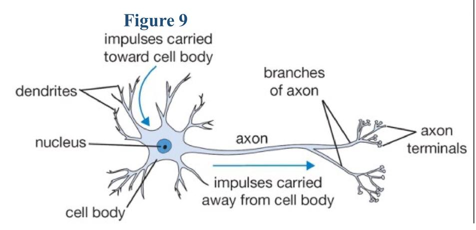
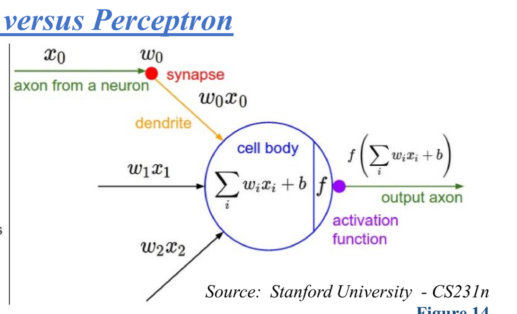


Figure 17

Biological Neuron versus Perceptron





Source: Stanford University - CS231n

Figure 14

- Biological neuron has dendrites to receive signals, cell body to process them, and an axon to send signals to other neurons
- At synapses between dendrite and axons, signals are modulated (increased/ decreased) in various amounts
- Neuron fires output signal when total input signal strength exceeds threshold

- Perceptron has several inputs.
- Inputs multiplied by weights assigned to that input.
- An offset (bias) is added to the sum of inputs. This is passed to a function whose output is the output of the perceptron

Model Accuracy

- **Figure 18** shows that model training had an AUC of 0.8570 and testing had an AUC of 0.8373
- **Figure 19** shows that 0.8006 is the average precision in testing

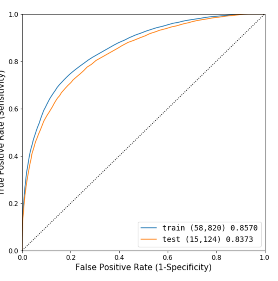


Figure 18

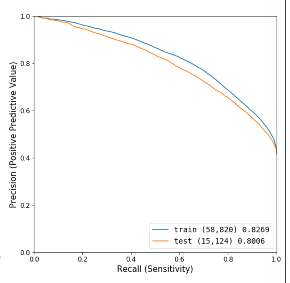


Figure 19

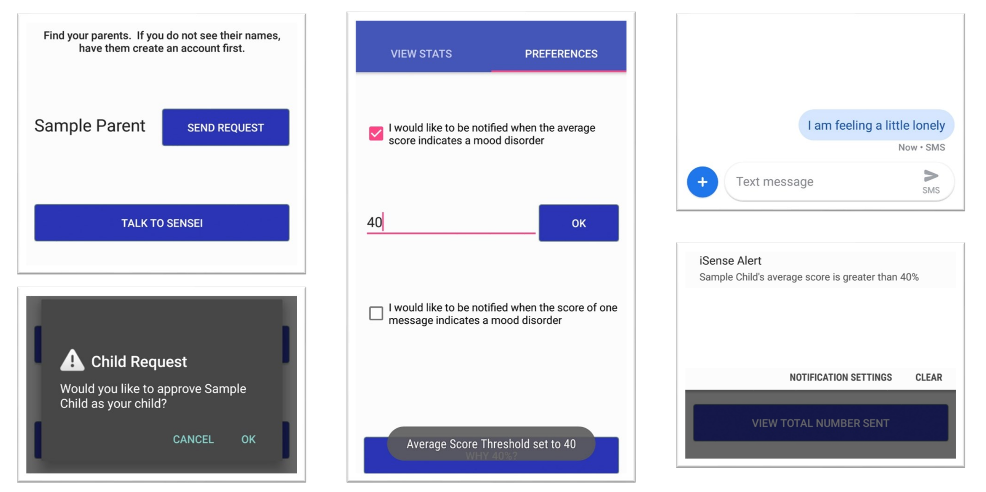
APP DESIGN/FUNCTIONS

- iSense is a fully-functional mobile app backed by an Artificial Intelligence model designed for a parent-teen audience

Teen and parent link accounts (after agreeing to terms of service)

Parent sets preferences for notifications

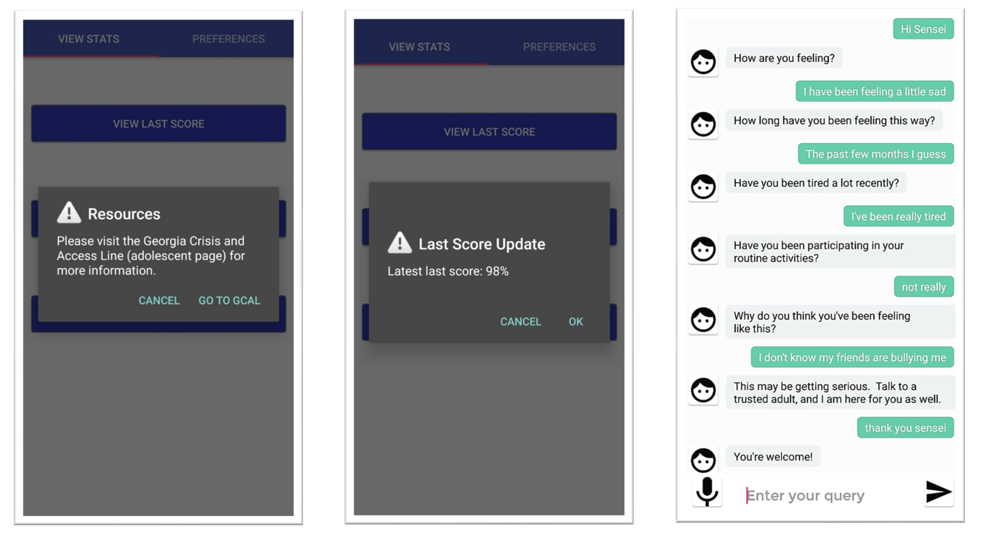
Teen sends message using SMS app, model scores message and saves score to database, alert sent based on score/preferences



Resources provided to parent when notification received

Real-time updates of most recent (last) score and average score

Voice-enabled chatbot in child interface, aligned with DSM-5 Criteria



PARAMETER OPTIMIZATION

GLM Lambda Search

- **Figure 20** shows how the number of coefficients increases as the objective function's lambda penalty is driven to zero
- **Figure 21** shows how as the number of coefficients increases, AUC increases
- 217 coefficients were selected in order to prevent overfitting, while maintaining an optimal AUC

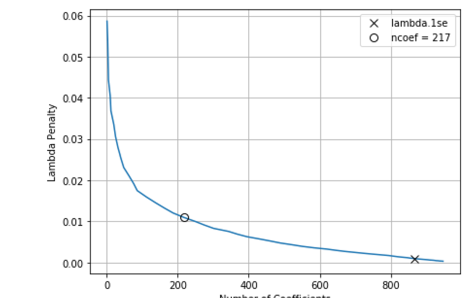


Figure 20

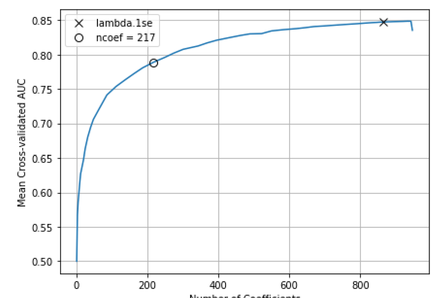


Figure 21

GBM Grid Search

- The table shows how Mean AUC and Binary Logloss Mean change as different combinations of parameters are tried (24 total combinations tried)
- The parameters displayed in the first row were used as the final parameters

	Learning Rate	Number of Leaves	Feature Fraction	Minimum Data/Leaf	Mean AUC	Binary Logloss Mean	Number of Trees
1	0.015	63	0.25	100	0.801	0.544	300
2	0.015	31	0.25	150	0.783	0.562	300
3	0.010	63	0.25	125	0.766	0.645	55
4	0.010	31	0.25	100	0.764	0.647	55
5	0.010	31	0.20	150	0.744	0.672	14

LINGUISTIC ANALYSIS

- As shown in **Figures 22 and 23,** all four of the tested linguistic patterns had significant odds ratios relevant risks
- Linguistic Patterns: indication of outward insecurity, enlarged use of indefinite pronouns, reference to drugs/withdrawal symptoms, enlarged use of self-referent words

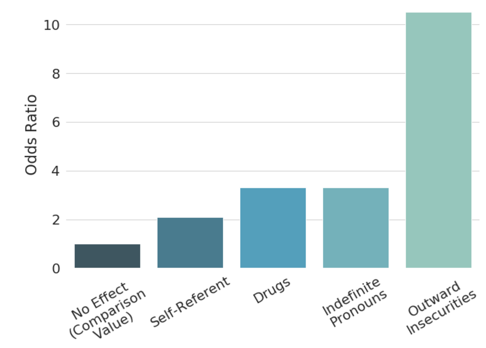


Figure 22

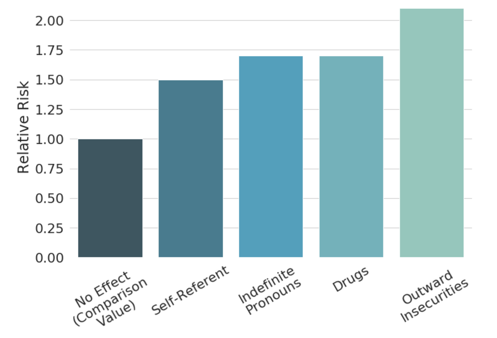


Figure 23

CONCLUSIONS

Conclusions

- The hypothesis was supported in that there exists a measurable difference in the language use of an individual with a mood disorder when compared to a healthy individual's language use
- After parameter optimization, the three models built resulted in the following F1 Scores:

	Generalized Linear Model		Gradient Boosting Machine		Neural Network	
	Train	Test	Train	Test	Train	Test
F1 Score	0.678	0.675	0.707	0.693	0.740	0.717

- 0.4 was determined to be the threshold because F1 Score is maximized at this value
- These conclusions were implemented in iSense, a mobile app that has the ability to to effectively and efficiently detect mood disorders

Future Work

- Future directions could include analyzing data from other social media sources to obtain a wider sample and a larger vocabulary
- The next iterations of the existing models could be trained to distinguish between existence of depression as part of a unipolar or bipolar disorder
- An iOS version of iSense could be created to allow for universal use