AN INTELLIGENT MOBILE APPLICATION TO IDENTIFY FACTORS INFLUENCING ADOLESCENT MENTAL HEALTH VARIABILITY USING ARTIFICIAL INTELLIGENCE

Richard Feng[1], Soroush Mirzaee[2]

[1]St. Margaret's Episcopal School, 31641 La Novia Ave,
San Juan Capistrano, CA 92675
[2]Computer Science Department, California State
Polytechnic University, Pomona, CA 91768

ABSTRACT

Artificial intelligence has shown promise in diagnosing mental illness in young children, a challenging task given the rise in teenagers struggling with mental health. We focus on the capabilities of machine learning and natural language processing models to accurately recognize activities that affect mental health in pre-teens and adolescents, an important step towards improving symptoms of depression and anxiety. We achieved an accuracy of 86.7% for determining sentiment from child journal entries with LSTM and BERT and a MSE of 94.6 for predicting future mental health outcomes with neural networks. We develop an innovative solution of incorporating these models inside of a mobile application as a scalable framework for data collection to track shifts in overall user wellbeing.

KEYWORDS

Artificial Intelligence, Sentiment analysis, Machine learning, Mental health

1. INTRODUCTION

In recent years, depression and anxiety in young children and adolescents have increasingly gained attention. Childhood depression is caused by genetic and social factors and has many long-term consequences that can affect adulthood well-being, including "suicidality, problems in social functioning, and poor physical and mental health" [1]. Adolescent mental illness is an urgent problem due to its prevalence and low treatment rates. It is estimated that 40% of children will meet the criteria for an identified mental health condition by the age of 18 [2]. In the last decade, diagnosis rates of childhood mental illness have increased dramatically. Owing to the complex nature of mental disorders, children are often underdiagnosed or misdiagnosed at the onset [11][12]. Moreover, the widespread nature of mental illness has put a severe strain on psychologists in the United States. In a 2023 survey, over half (56%) of the psychologists have no new openings, and more than a third are "burnt out" [3].

A novel approach involves using mobile applications and artificial intelligence (AI) as a cost-effective and widely available support tool [4]. Previous research demonstrates the efficacy of incorporating AI and natural language processing (NLP) into mobile applications to aid in

David C. Wyld et al. (Eds): IoTE, CNDC, DSA, AIAA, NLPTA, DPPR - 2024
pp. 95-102, 2024. - CS & IT - CSCP 2024 DOI: 10.5121/csit.2024.142308

identifying mental health conditions and providing user support [5]. This paper proposes a mobile application leveraging machine learning (ML) and NLP to predict mental health changes by identifying factors that influence mental health, providing a scalable and personalized framework for early intervention.

2. Solution

Different from social studies, we formulate the problem of mental illness as a statistical and machine learning question that identifies changes in mental health to further identify possible causes, an approach that leverages data-driven techniques including regression, tree-based methods, and NLP [6]. Our study has three main components:

Data generation: Collecting journal entries and daily activity logs from users.

Sentiment analysis: Analyzing user inputs to determine emotional tone and mental state.

Future mental health prediction: Using ML models to forecast mental health outcomes.

An important part of mental illness diagnosis is recognizing changes in mental health and fluctuations in mood early on. To effectively train our ML model to identify causation in mental health changes over time, an extensive dataset is necessary. We explore the use of large language models (LLMs) in generating experimental data, a creative solution that avoids the challenges of recruiting a large user base. We write custom instructions for an LLM to generate a set of data to replicate the user behavior and journal entries in the application. Then, the output data is run through the Sentiment Analysis Model that is trained to rate the sentiment of the entries to calculate a mental health score. The scores and user application data are fed into the User Activity Model to analyze how daily activities influence user well-being.

2.1. User and Journal Data

We devise two custom GPTs using OpenAI's API to create users and journal entries. The first generates users profiles with the following details: ages between 8-20 following a standard distribution, random genders, race/ethnicity and city location based on the proportions of the US population, random socioeconomic status, parental involvement, hobbies, and random personality type according to the Big Five Personality Type. These instructions ensure the diversity of the users and variance in the journals generated. We generate users in 12 batches of 25 to achieve optimal results to avoid loss of attention in the thread for a total of 300 users. The second custom GPT generates journal entries given the profile of each child. We task the agent to generate 5 journal entries before the study; randomly decide if the user's mental health improves, stays the same, or worsens; and 5 journal entries after the study that takes into account the child's profile.

2.2. Sentiment Analysis

We create five different models for sentiment analysis: Logistic Regression, RandomFores, Naive Bayes, Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT) [14][15][16][19]. The models are trained on a dataset of 1.6 million tweets labeled as positive (1) or negative (0) [9]. The steps of training each model are as follows:

Data Preprocessing: Cleaning and preparing the data for analysis.

Training: Building and training the machine learning models on the processed data.

Hyperparameter Optimization: Fine-tuning the models to enhance performance.

Our model calculates the mental health score of a journal entry as a percentage, which serves as a critical indicator of a user's wellbeing over time. For each user, the scores are aggregated across all journal entries labeled as before and after to calculate a before and after score respectively.

2.3. User Activity Data

From the mental health scores, the user interaction data is generated using a rule-based algorithm. For example, positive activities such as exercise and reading are logged more frequently for users with improving mental health scores, reflecting evidence from psychological studies on the benefits of these activities. Specifically, our data simulates the number of times a user logs a certain activity in the app in the timeframe of 6 months (i.e., 180 days). First, the activities are categorized by their impact and correlation on mental health based on existing research. We consider the features rainy, windy, movie & TV [26], gaming [23], homework, and exam [28] as negative; cloudy, snowy, class, and study as neutral; journals [10][18], sunny [25], reading [32], walking, and drawing [33] as slightly positive; instrument and sleep [31] as moderately positive; and gratitude [27], exercise [29], and music [30] as highly positive. For each category, a different generation rule is applied that considers the change in the mental health score of the users. For positive categories, a random, large number is generated for users with an increase in mental health and vice versa. Special rules are used for journaling and sleep. The randomness ensures that the data will not be linear and have outliers, but still retain a slight trend.

2.4. User Activity Model

We investigate the relationship between daily activities and mental health outcomes, focusing on the impact of these activities on the Mental Health Score. Our dataset consists of user data, with each row representing an individual user and columns with the mental health score before and after the observed period and the activities listed above. We use a suite of five ML models for this task: Linear Regression, Random Forest, XGBoost, Support Vector Machine (SVM), and Neural Networks. We start by preprocessing the data by handling missing values through imputation methods and normalizing the features to ensure they are on a comparable scale. Then, we train each model and perform feature selection.

3. EXPERIMENT

This section delves into the experimental methodologies and analytical techniques applied in our study to assess and predict mental health outcomes. We focus on two main areas: sentiment analysis and user activity. In the sentiment analysis component, we preprocess journal data using various techniques and compare the performance of multiple ML models. In the user activity component, we systematically identify and evaluate features that influence mental health scores, using scientific model development and evaluation metrics to ensure accuracy and robustness in our predictions.

3.1. Sentiment Analysis

We preprocess the journal data using tokenization, stopword removal, and TF-IDF vectorization [19][20][21]. For deep learning models, we use word embeddings such as Word2Vecto capture semantic relationships [22]. We compare the performance of Logistic Regression, Random Forest, Naive Bayes Classifier, LSTM, and BERT in analyzing sentiment and perform hyperparameter optimization on the models with RandomSearch.

Table 1. Sentiment Analysis Model Accuracy

Model	Model 1	Model 2	Model 3	Model 4	Model 5
Data Process	TF-IDF Vectorization	TF-IDF Vectorization	NLTK	Gensim	transformers
Model	Logistic Regression	Random Forest	NaiveBayesClassifier	LSTM	Bert
Train Accuracy	0.77218	0.96697	0.79776	0.81326	0.90322
Validation Accuracy	0.77302	0.75501	0.75213	0.78017	0.88357
Test Accuracy	0.76939	0.75333	0.75758	0.78322	0.86752

Model 1: TfidfVectorizer + Logistic Regression

We use TfidfVectorizer to transform the text data into numerical features based on term frequency-inverse document frequency (TF-IDF), capturing the importance of words in the dataset. We also feed numerical features from vectorized into a Multinomial Naïve Bayes classifier assuming that word occurrences follow a multinomial distribution.

Model 2: TfidfVectorizer + Random Forest

The Random Forest model, also using TF-IDF vectorization, exhibited the highest train accuracy of 96.69%, a sign of potential overfitting. Its validation and test accuracies were lower than other models, suggesting that while it fits the training data, it may not generalize as effectively to unseen data.

Model 3: WordNetLemmatizer + Naive Bayes Classifier

We use WordNetLemmatizer to preprocess text by reducing words to their base or root forms. A Naive Bayesian Model, which relies on Bayes' theorem to predict the probability of each sentiment class, uses the lemmatized text to provide an effective method for sentiment analysis.

Model 4: Gensim Word2Vec + LSTM

Gensim Word2Vec is used to create word embeddings, capturing semantic relationships between words in a continuous vector space. These embeddings are fed into an LSTM network, which includes layers for embedding, dropout to prevent overfitting, LSTM for capturing sequential dependencies, and a Dense layer for final predictions.

Model 5: BERT using CUDA

The BERT model is trained using CUDA for efficient computation on Nvidia GPUs. BERT has bidirectional transformers to understand the context of each word in the entire sentence.

Out of these five models, BERT produced the highest train accuracy of 86.7%. Despite this, we select LSTM instead as BERT tends to produce extremely polar (>99 and <1) probabilities.

3.2. User Activity

We build the User Activity Model with the following procedures:

Feature Engineering

We use a systematic and iterative approach to identify the most relevant features for predicting the mental health score. First, we include a broad set of features representing various daily activities. We calculate the Pearson correlation coefficients between each feature and the after mental health score. Features with high correlation coefficients are considered strong candidates for inclusion in the model, as they showed a significant linear relationship with the target variable. Features that did not significantly improve the model's performance are removed. To check for multicollinearity between the features, we calculate the Variance Inflation Factor (VIF) for each feature. We use feature importance metrics from more complex models like Random Forest and XGBoost as an additional criterion.

Model Development

We use the ordinary least squares method to minimize the sum of squared residuals, providing coefficients for each feature that indicate their linear relationship with the after mental health score. To ensure optimal performance, we apply GridSearchCV to RandomForest and SVM models to systematically explore the hyperparameter space and identify the best combination of parameters. We use RandomizedSearchCV for XGBoost and Neural Network models to efficiently search a broad hyperparameter space with random sampling. We also perform k-fold cross-validation to evaluate the performance and generalizability of each model.

Evaluation Metrics

The performance of the models is assessed using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and the R^2 Score. As shown in Table 3, the Neural Network demonstrated higher performance with the lowest Testing MSE of 94.560, indicating the highest accuracy in predicting mental health scores.

Table 2. User Activity Model Accuracy

Model	Model 1	Model 2	Model 3	Model 4	Model 5
Model	Linear Regression	Random Forest	XGBoost	SVM	Neural Network
Testing MSE	164.974	113.830	108.310	571.690	94.560
Testing MAE	0.727	0.810	0.820	0.050	0.840
Testing R2 Score	10.447	7.850	7.383	15.830	7.286

The importance scores provided by the logistic regression and random forest models are similar, allowing us to conclude that specific features have a greater impact on the target variable. For example, exercise, music, and gratitude show high importance in both models, suggesting that these activities positively affect mental health, whereas exam, rainy and gaming have negative impacts. Our models did not incorrectly classify features designated positive earlier as negative and vice versa. The only instance when these models showed discrepancy is classifying neutral features incorrectly, or classifying slightly positive or negative features as neutral. Our models show accuracy in identifying feature weights, indicating that they can understand the positive and negative influences on mental health to optimize interventions for enhancing well-being.

4. RELATED WORK

Detecting Mental Illness

The use of ML and deep learning in detecting mental illness has long been studied. One common scenario is using data from Twitter and Reddit to identify depression and suicidality [6]. A majority of studies use supervised learning, which involves labeling a large amount of data. Among the methods used are SVM, k-nearest Neighbors (KNN), Decision Tree, Random Forest, Naive Bayes (NB), Logistic Regression, and XGBoost[7]. We expand on this work by creating a scoring system used as a comparison for long-term changes in emotional state.

AI and Adolescent Mental Health

Previous research has pointed to the potential of machine learning in addressing youth mental illness. One 2020 study used several ML models to analyze mid-teen survey data with reasonable accuracy, but not enough for clinical use [8]. Online mental health surveys that are not recognized can often be misleading and false, hence we analyze journals instead.

5. CONCLUSIONS

From our experimental results and analyses, there are several limitations and areas of improvement in future works.

The Sentiment Analysis Model is trained on a dataset that lacks clinical validity, as it uses Tweets labeled as "positive" and "negative," rather than clinically relevant categories. Models trained on text written by children instead of adults with such a classification can achieve more accurate results for this study. While we achieve an accuracy of 86.7%, the accuracy of our predicted scores cannot be verified beyond whether it is overall positive or negative because of the dataset's binary classification.

In addition, the simplicity of the rule-based algorithm used to generate user activities does not account for real-world user behavior. These activities differ from user to user. Moreover, the randomness in the generation means that there is no direct correlation between different features that could exist in real life. For instance, we may see an inverse relationship between sleep time and exams.

Our study suggests that machine learning can be used in mobile applications to predict changes in child mental health based on written journal entries and activities, a unique perspective of viewing the problem of adolescent depression by finding causation in data. We achieve our best results with LSTM for determining journal sentiment and neural networks for predicting well-being outcomes. This framework can reduce the burden on mental health professionals and help individuals and families better manage mental health, potentially reducing the long-term effects of untreated conditions. Future work should focus on validating these models with real-world data and integrating such tools into existing healthcare systems for maximum benefit.

REFERENCES

[1] Maughan, Barbara, Stephan Collishaw, and Argyris Stringaris. "Depression in childhood and adolescence." Journal of the Canadian Academy of Child and Adolescent Psychiatry 22.1 (2013): 35.

[2] Shim, Ruth, Moira Szilagyi, and James M. Perrin. "Epidemic rates of child and adolescent mental health disorders require an urgent response." Pediatrics 149.5 (2022).

[3] American Psychological Association. "Psychologists reaching their limits as patients present with worsening symptoms year after year." 2023,

[4] Chandrashekar, Pooja. "Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps." Mhealth 4 (2018).

[5] Milne-Ives, Madison, et al. "Artificial intelligence and machine learning in mobile apps for mental health: A scoping review." PLOS digital health 1.8 (2022): e0000079.

[6] Zhang, Tianlin, et al. "Natural language processing applied to mental illness detection: a narrative review." NPJ digital medicine 5.1 (2022): 1-13.

[7] Kumar, Piyush, et al. "A machine learning implementation for mental health care. Application: smart watch for depression detection." 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2021.

[8] Tate, Ashley E., et al. "Predicting mental health problems in adolescence using machine learning techniques." PloS one 15.4 (2020): e0230389.

[9] Tan, Kian Long, Chin Poo Lee, and Kian Ming Lim. "A survey of sentiment analysis: Approaches, datasets, and future research." Applied Sciences 13.7 (2023): 4550.

[10] Sohal, Monika, et al. "Efficacy of journaling in the management of mental illness: a systematic review and meta-analysis." Family medicine and community health 10.1 (2022).

[11] Tkacz, Joseph, and Brenna L. Brady. "Increasing rate of diagnosed childhood mental illness in the United States: Incidence, prevalence and costs." Public Health in Practice 2 (2021): 100204.

[12] Reimherr, J. Philip, and Jon M. McClellan. "Diagnostic challenges in children and adolescents with psychotic disorders." Journal of Clinical Psychiatry 65 (2004): 5-11.

[13] Clark, Molly S., Kate L. Jansen, and J. Anthony Cloy. "Treatment of childhood and adolescent depression." American family physician 86.5 (2012): 442-448.

[14] Samanvitha, S., et al. "Naïve Bayes Classifier for depression detection using text data." 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT). IEEE, 2021.

[15] Apoorva, Abhyudaya, et al. "Depression detection on twitter using RNN and LSTM models." International Conference on Advanced Network Technologies and Intelligent Computing. Cham: Springer Nature Switzerland, 2022.

[16] Kumari, Meena, Gurpreet Singh, and Sagar Dhanraj Pande. "Depressonify: BERT a deep learning approach of detection of depression." EAI Endorsed Transactions on Pervasive Health and Technology 10 (2024).

[17] Wurtz, Heather M., Sarah S. Willen, and Katherine A. Mason. "Introduction: Journaling and mental health during COVID-19: Insights from the pandemic journaling project." SSM-Mental Health 2 (2022): 100141.

[18] Smyth, Joshua M., et al. "Online positive affect journaling in the improvement of mental distress and well-being in general medical patients with elevated anxiety symptoms: A preliminary randomized controlled trial." JMIR mental health 5.4 (2018): e11290.

[19] Kholifah, Binti, Iwan Syarif, and Tessy Badriyah. "Mental disorder detection via social media mining using deep learning." Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control 5.4 (2020): 309-316.

[20] Jayapal, Cynthia, et al. "Detection of Mental Health Using Deep Learning Technique." International Conference on Communication and Intelligent Systems. Singapore: Springer Nature Singapore, 2022.

[21] Shetty, Nisha P., et al. "Predicting depression using deep learning and ensemble algorithms on raw twitter data." International Journal of Electrical and Computer Engineering 10.4 (2020): 3751.

[22] Rahman, Tazrin, et al. "Text-Based Data Analysis for Mental Health Using Explainable AI and Deep Learning." Networking and Parallel/Distributed Computing Systems: Volume 18. Cham: Springer Nature Switzerland, 2024. 17-31.

[23] Von der Heiden, Juliane M., et al. "The association between video gaming and psychological functioning." Frontiers in psychology 10 (2019): 1731.

[24] Beecher, Mark E., et al. "Sunshine on my shoulders: Weather, pollution, and emotional distress." Journal of affective disorders 205 (2016): 234-238.

[25] McAnally, Helena M., Tamara Young, and Robert J. Hancox. "Childhood and adolescent television viewing and internalising disorders in adulthood." Preventive medicine reports 15 (2019): 100890.

[26] Zhao, Liange, Hongbin Yuan, and Xueyuan Wang. "Impact of homework time on adolescent mental health: Evidence from China." International Journal of Educational Development 107 (2024): 103051.

[27] Emmons, Robert A., and Michael E. McCullough. "Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life." Journal of personality and social psychology 84.2 (2003): 377.

[28] Fritz, Jessica, et al. "Tracking stress, mental health, and resilience factors in medical students before, during, and after a stress-inducing exam period: Protocol and proof-of-principle analyses for the RESIST cohort study." JMIR Formative Research 5.6 (2021): e20128.

[29] Sharma, Ashish, Vishal Madaan, and Frederick D. Petty. "Exercise for mental health." Primary care companion to the Journal of clinical psychiatry 8.2 (2006): 106.

[30] Rebecchini, Lavinia. "Music, mental health, and immunity." Brain, behavior, & immunity-health 18 (2021): 100374.

[31] Short, Michelle A., Kate Bartel, and Mary A. Carskadon. "Sleep and mental health in children and adolescents." Sleep and health. Academic Press, 2019. 435-445.

[32] Dowrick, Christopher, et al. "Get into Reading as an intervention for common mental health problems: exploring catalysts for change." Medical Humanities 38.1 (2012): 15-20.

[33] Heenan, Deirdre. "Art as therapy: an effective way of promoting positive mental health?." Disability & society 21.2 (2006): 179-191.