Automated ADHD Symptom Identification and Integration into Zenbo Chatbot

Student: 郭俐欣 B09902083
Advisor: 傅立成 教授

Motivation

Attention deficit hyperactivity disorder (ADHD) is a condition that affects people's behavior. People with ADHD can seem restless, may have trouble concentrating and may act on impulse. ADHD is classically seen as a childhood disease, although it persists in one out of two cases in adults. The diagnosis is based on a long and multidisciplinary process, involving different health professionals, leading to an underdiagnosis and undertreatment of adult ADHD. The significance of identifying and addressing ADHD is on the rise, given its widespread recognition as a condition that hinders an individual's ability to interact effectively with their environment, thereby elevating the potential for harmful events.

Project Overview

The aim is to develop a text classification system for automatically identifying ADHD symptoms in text responses. Followed by an integration of this system into a chatbot that runs on Zenbo, a physical robot developed by Asus. The chatbot will facilitate real-time interactions with users, providing a user-friendly interface for ADHD symptom assessment, support, and information dissemination.

Previous Work and Datasets

A Classified Mental Health Disorder (ADHD) Dataset Based on Ensemble Machine Learning from Social Media Platforms (Sabrina Mostafij Mumu, Hasibul Hoque & Nazmus Sakib):

In this study, machine learning is used as the baseline to diagnose the mental health condition known as Attention Deficit Hyperactivity Disorder (ADHD). An unsupervised dataset for ADHD was converted into a classified dataset. To get better results, an ensemble machine learning model made up of five classifiers was deployed. Linguistic models were employed to improve the machine's ability to predict. This recent dataset offers the potential for future developments such as chatbots that can be developed to provide people with psychological support and the application of natural language processing to identify the behavior pattern of ADHD patients.

The corpus of this dataset, titled "Classified ADHD Dataset", consists of two columns named sentence and label. Each column contains around eighteen thousand grades. The sentence column contains text data from the previously mentioned datasets. Highly scored data from "ADHD", "adhdwomen" and positive emotion texts from "Emotions in Text" were collected. The label column indicates whether or not the corresponding sentence reflects ADHD. Tables 1-3 present the datasets used to create the "Classified ADHD Dataset" (table 4).

Table 1 ADHD dataset

Title	Selftext	Score	Id	Created-utc	Created-datetime
Not feeling affects of aderall.	I've been on aderall for a couple weeks now and it seems like the boost in my ability to pay attention has already faded	30	kzux1	1239041887	2009-05-31 17:08:19

Table 2 ADHD women dataset

Body	Score	id	Created-utc	Created-datetime
You definitely need to talk to HR. Politely. It's illegal for them to deny you a job for a legitimate medical condition and prescribed medications	345	cy5o5j8	1304183283	2011-04-30 17:08:03

Table 3 Positive sentiment dataset

Text	Emotion
I am ever feeling nostalgic about the fireplace i will know that it is still on the property	Love
I have been with petronas for years i feel that petronas has performed well and made a huge profit	Happy

Table 4 Classified ADHD dataset

Sentence	Label
I feel like all my struggles in the past and present make sense now. I always thought I was too dumb or lazy to do well in school, because it was so hard to focus on anything. I've dropped out of college twice. First time due to failing, second time I impulsively withdrew from all my classes the night before the semester started. Now that I know I have ADHD I am determined to learn how to manage it, so that I can return to school. Anyone else a college student? How do you manage? I have problems with studying for more than five minutes at a time	Yes
I enjoyed this semester and i enjoyed the challenges i got to face and overcome and i feel that i m really coming away with a lot of valuable experience out of this	No

Questionnaire-based computational screening of adult ADHD(Arthur Trognon & Manon Richard):

The DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, 5th edition) is a widely used diagnostic manual in the field of mental health. It provides criteria for the diagnosis of various mental disorders, including ADHD(Attention-Deficit/Hyperactivity Disorder). The DSM-5 criteria for ADHD were used as the basis for developing the questionnaire-based screening scale for adult ADHD.

In the study, the researchers designed a scale consisting of 43 preliminary items based on the DSM-5 criteria. This preliminary scale, called pTRAQ (Preliminary Trognon & Richard ADHD Questionnaire), was administered to a group of ADHD individuals and a control group. Statistical analysis was then conducted to select the most discriminating items that could differentiate between the two groups.

Based on the results of the analysis, the researchers generated the final version of the questionnaire, called TRAQ10 (Trognon & Richard ADHD Questionnaire French 10-items). This scale consisted of 10 items that were considered to be the most relevant and effective in identifying adult ADHD. The following is a chart describing these 10 items.

Project ID	Screener ID	Factor	Points	English Item (Author's suggestion)
pTRAQ3	TRAQ1	Attention	6	I often make careless mistakes in my activities.
pTRAQ11	TRAQ2	Attention	6	I have difficulty staying focused during my activities (domestic, professional).
pTRAQ41	TRAQ3	Impulsivity	6	It is difficult for me to wait my turn in a queue.
pTRAQ4	TRAQ4	Attention	6	I have trouble maintaining my attention at work.
pTRAQ1	TRAQ5	Impulsivity	6	I don't really pay attention to details.
pTRAQ28	TRAQ6	Impulsivity	6	I often leave my seat unnecessarily during a meeting.
pTRAQ26	TRAQ7	Impulsivity	6	I often wiggle my hands or feet on my seat
pTRAQ24	TRAQ8	Attention	6	I am often subject to forgetfulness in my daily life (doing housework, shopping...)
pTRAQ14	TRAQ9	Impulsivity	6	It is difficult for me to organize tasks that require several steps
pTRAQ7	TRAQ10	Attention	6	My relatives blame me for not listening to them when they talk

Model Development

Preprocessing

The original data set was split into 80% train data and 20% test data. The labels "Yes" and "No" were encoded as 1 and 0 respectively.

Data Cleaning using Natural Language Toolkit (NLTK):

Natural Language Toolkit (NLTK) is an open source Python library for Natural Language Processing. It has functionality for almost all text processing techniques and a suite of text processing libraries like classification, tokenization, stemming, tagging, parsing, and semantic reasoning, lemmatization, POS Tagging and also wrappers for industrial-strength NLP libraries.

This library was used to write a series of functions to help clean the dataset by removing urls, html tags, lowercase the text, numbers, punctuations, stop words, mentions, hashes, and spaces. Another function was also written to perform stemming on the data set. Stemming is a text normalization technique in natural language processing (NLP) and information retrieval that involves reducing words to their base or root form. The root form is often the word stem, which may or may not be a valid word. Also, a function was also written to perform lemmatization which aims to reduce words to their base or dictionary form (known as the lemma) by considering the context and meaning of the word. The resulting lemmas are usually valid words.

Extracting Features from the Text

The Bag of Words (BoW) representation is a common technique in NLP for converting text data into numerical vectors. The BoW model represents a document as an unordered set of words and their frequencies. The Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer is a variation of the BoW model that takes into account the importance of words in a document relative to their occurrence across multiple documents.

After achieving a cleaned set of records, we need to represent text with features that models can understand. This was done with BoW representation using TF-IDF vectorizer in sklearn and by applying singular value decomposition (SVD) on top of it to efficiently reduce the dimensionality before sending it to models.

Running ML Algorithms

In this project, the initial step involved working with a feature matrix and applying various popular algorithms known for their effectiveness in text classification. The SVD representation of BoW features was utilized for this purpose. Specifically, three distinct techniques were chosen: linear support vector machines, logistic regression, and a random forest classifier from the sklearn library. This selection allowed for an exploration of different foundational assumptions in modeling, facilitating a comprehensive comparison. In this project, the cross-validation process was executed to efficiently evaluate the model. For K-fold cross-validation (CV), a choice of 5 folds was made. Utilizing k-fold cross-validation

enables an assessment of the model's performance across various cuts of the dataset, as opposed to relying on a single cut alone.

Comparing Accuracy and Plotting Model Results

The following chart illustrates the scores of the three models:

	Model	Score
0	SVD_BoW_SVM	0.942688
1	SVD_BoW_LogisticRegression	0.940994
2	SVD_BoW_RandomForest	0.951158

As we can see from the chart, Random Forest Classifier was the model that had the best performance, with a score of 0.951158. Plotting of the confusion matrices for each model had the following results:
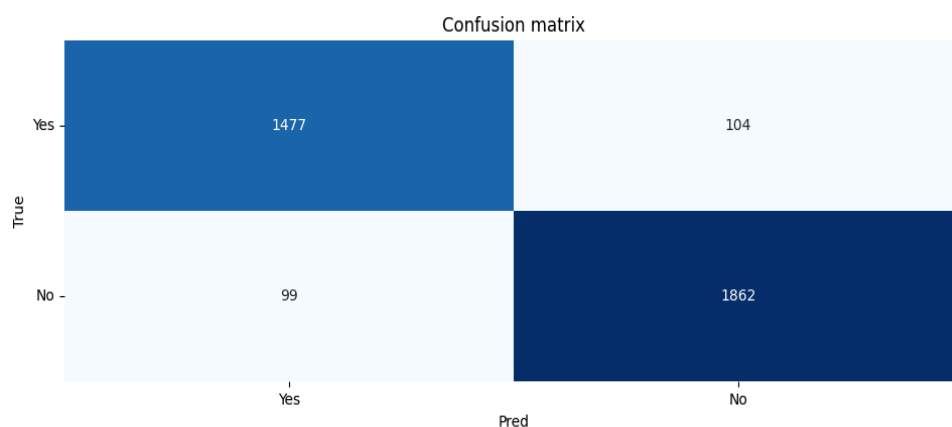
Random Forest Classifier:
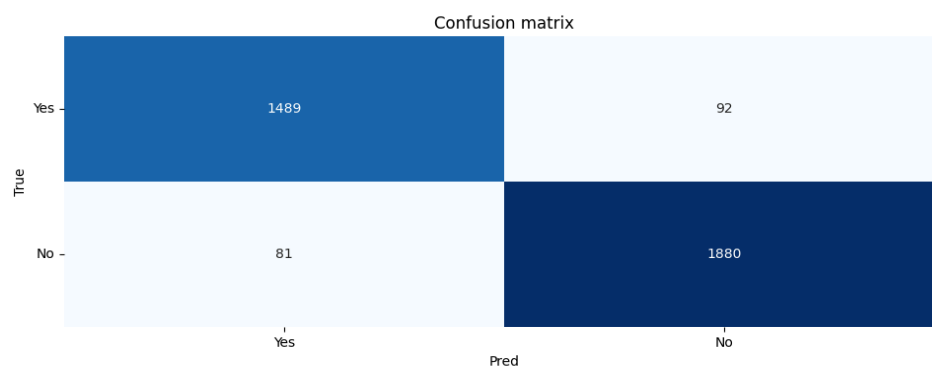
Logistic Regression:
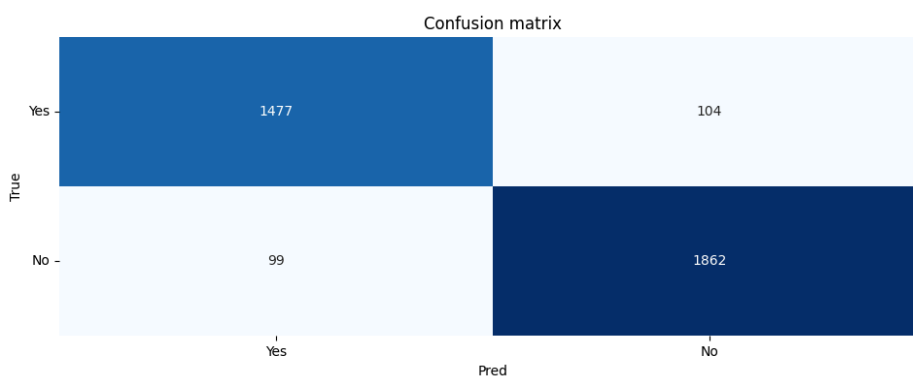
Linear SVM:

Thus, in order to train a final model, Random Forest Classifier, would be the ML algorithm of choice due to its prominent classification performance in comparison to the other ML algorithms.

Questionnaire Development

The completion of a daily questionnaire can help an individual identify behaviors associated with attention and focus challenges can be helpful for tracking patterns over time. The following questionnaire has 12 items. The first 10 items are based on the TRAQ10, a questionnaire consisting of 10 items that were considered to be the most relevant and effective in identifying adult ADHD. The final two items aim to grasp the individual's general emotional state:

1. **Observations in Activities:** Can you share any observations you made during your activities today?

2. **Engagement in Domestic Activities:** What aspects of your domestic activities held your attention today?

3. **Waiting in Queues:** How did you spend your time while waiting in a queue today?

4. **Work Highlights:** Reflect on different aspects of your work today. What tasks or moments stood out to you?

5. **Attention to Details:** Recall instances where you noticed details in your tasks today. What specific things caught your attention?

6. **Participation in Meetings:** Describe your level of engagement during meetings today. Were there any notable aspects?

7. **Comfort Movements:** Were there any comfortable or natural movements, like stretching or adjusting, during your seated activities today?

8. **Daily Task Recall:** Can you recall how you approached daily tasks, such as housework or shopping, today?

9. **Task Organization:** Share your observations on organizing tasks with multiple steps today.

10. **Communication with Relatives or Friends:** Reflect on your communication with relatives or friends today. What stood out during your conversations?

11. Are there any other important or relevant events or details from your day you would like to reflect on?

12. Overall, how do you feel about how your day unfolded?

Outstanding Tasks

Integration to Zenbo

After the successful training of our final machine learning model, the next phase involves integration into the Zenbo platform. This integration would be accomplished by expanding upon the features of the existing mental health app the lab has developed. This innovative addition aims to leverage the newly trained model to enhance the app's capabilities in identifying and assisting users with ADHD-related challenges. The chat feature will facilitate a daily questionnaire, engaging users in a conversational manner and intelligently keeping track of their responses. These responses will be later classified by the learning model, contributing to a more nuanced understanding of individual patterns and aiding in ADHD diagnosis. By seamlessly incorporating advanced machine learning techniques into the familiar interface of the Zenbo platform, our goal is to offer a tailored and user-friendly experience that provides valuable insights for individuals seeking ADHD diagnosis and support.

Testing

Another outstanding task in this project involves conducting thorough testing with real individuals. This crucial step would aim to gather valuable user feedback and insights, ensuring the optimal performance. During this testing phase, a key focus will be on determining the ideal timeframe for users to respond to the daily questionnaire. By actively involving real users, we aim to understand their preferences and daily routines, ultimately refining the user experience and ensuring that the app aligns seamlessly with their lifestyles.

Conclusions

By integrating the ADHD symptom identification model into a chatbot on Zenbo, this project aims to create an innovative tool for early symptom recognition, support, and information dissemination related to ADHD. The combination of NLP techniques, robotics, and real-time user interaction opens new possibilities for assisting individuals and families affected by ADHD.

References

- Mumu, S.M., Hoque, H., Sakib, N. (2023). A Classified Mental Health Disorder (ADHD) Dataset Based on Ensemble Machine Learning from Social Media Platforms. In: Kaiser, M.S., Waheed, S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds) Proceedings of the Fourth International Conference on Trends in Computational and Cognitive Engineering. Lecture Notes in Networks and Systems, vol 618. Springer, Singapore. https://doi.org/10.1007/978-981-19-9483-8_33
- Trognon, A., Richard, M. Questionnaire-based computational screening of adult ADHD. *BMC Psychiatry* 22, 401 (2022). https://doi.org/10.1186/s12888-022-04048-1

- https://www.frontiersin.org/articles/10.3389/fninf.2020.575999/fullFrontiers | Machine Learning Methods for Diagnosing Autism Spectrum Disorder and Attention-Deficit/Hyperactivity Disorder Using Functional and Structural MRI: A Survey

- https://link.springer.com/article/10.1007/s10439-023-03269-z

- https://www.nhs.uk/conditions/attention-deficit-hyperactivity-disorder-adhd/

- https://www.scaler.com/topics/nlp/nlp-text-classification-using-scikit-learn/