

MAPI Chatbot: NLP Application For Physical and Mental Illnesses Identification

Tusheet Sidharth Goli

Tejas Pradeep

Akshay Sathiya

Pranav Khorana

Rahul Chawla

Sanket Manjesh

College of Computing, Georgia Institute of Technology
{tgoli3, tpradeep8, asathiya6, pkhorana3, rchawla36, smanjesh3}@gatech.edu

Abstract

A primary motivation behind this project is the increasing prevalence of mental health issues in recent years, in addition to ongoing physical health issues. Researchers estimate the number of adults with some degree of serious psychological distress has increased by 71% from 2006 to 2017, among young adults (Rosenberg, 2019). Our goal is to implement a project that would be a chatbot system resembling a social media interface that can use machine learning to predict if the user is at risk of any mental illnesses or physical illnesses. The features we implement include an Interactive UI for a mobile application, a machine learning model that uses natural language processing to analyze user symptoms and recommend possible illnesses, and a database to store messages and responses for better future training data.

Introduction

Artificial Intelligence and Machine Learning are some of the most commonly used and most rapidly developing technologies today. These areas of technology are prevalent in almost every field today, including the field of healthcare. The advent of Machine Learning within the field of healthcare can offer a multitude of opportunities for self-care where patients can easily receive accurate diagnoses in a very convenient manner from their own home. With the ongoing prevalence and attention brought forward to mental health illnesses along with the constant prevalence of physical illnesses, it is imperative to consider methods that can help diagnose patients in a much quicker and more efficient manner than typical doctor checkups. Therefore, this paper looks into developing a Machine-Learning based chatbot to help patients self-diagnose various physical and mental illnesses and help them save time and money.

Background and Significance

Since Machine Learning is such a rapidly developing and commonly utilized field within society, it is very useful to try and test out its capabilities within the healthcare field. The usage of Machine Learning within the healthcare field can help speed up manual processes done by doctors and

other healthcare professionals and offer diagnoses and results that are as accurate or even more accurate than those given by healthcare professionals. Therefore, with the advent of mental health issues and physical illnesses within today's society, it is quite reasonable to try and integrate Machine Learning models for faster diagnoses of these illnesses and more convenient usage that can save users time and money.

The approach this paper attempts to take is to utilize a Machine-Learning based chatbot where users can type in messages and get back responses that can classify a user's condition based on a physical or mental illness they might have. The major significance of this method is that users can conveniently self-diagnose themselves from wherever they are located instead of having to spend time and money to schedule and go to a doctor's appointment. In addition, patients around the world who might not have the financial resources to access healthcare professionals can more easily get diagnoses for their physical and mental conditions and can find ways to treat these illnesses that would have been impossible otherwise. This chatbot can also be integrated with healthcare professionals and their diagnoses to help these professionals diagnose their patients on a more individual basis based on the messages they send and the illness classifications added by the healthcare professionals.

Problem

Underdiagnosis of diseases is a problem with current health infrastructure that prevents many from receiving the care they need to live healthy day-to-day lives. The World Health Organization (WHO) claims that less than half of those who meet criteria for mental health illnesses are actually identified and diagnosed by doctors [3]. Many believe that this could be caused by a stigma or illegitimate perception of several mental health illnesses. Thus, accessibility to quality healthcare and education about symptoms pertaining to mental health illnesses may be hard to find for several people.

These issues are not exclusive to mental illnesses, however. Even with physical illnesses, if patients do not have the means to visit a doctor or communicate their

symptoms clearly for a professional, they may have an illness that goes undiagnosed, which can cause more harm over time. Overall, for both categories, it is found that not communicating symptoms, or overall lack of access to quality health care can prevent diagnosis of illnesses, and in turn prevent individuals from obtaining access to quality care.

Solution

The MAPI chatbot aims to provide a unique approach to self-care and self-diagnosis. The goal of the chatbot is to provide a preliminary diagnosis for mental and physical illnesses based on user-reported symptoms. The aim is to provide users with insight into the possible seriousness of their symptoms and provide additional facts to help users make the decision to visit a doctor. The aim of the chatbot is to seem natural and use a predictive model to predict illness based on user symptoms. The UI of the chatbot is interactive and resembles a social media platform helping the user feel more comfortable while mentioning their symptoms. Further to facilitate natural conversation, the chatbot uses Natural Language Processing to extract symptoms from the conversation and further extract the emotions of the user messages to facilitate the mental illness predictor. Currently, the chatbot can diagnose a wide variety of physical illnesses but is limited to diagnosing three mental illnesses namely depression, anxiety, and bipolar disorder. To determine what mental illness the user has, the chatbot has a set of predefined heuristics for each of the mental illnesses. To classify a user as having depression, the majority classification for each segment of the message sent by the user has to be under the classification of “sadness”. To classify a user as having anxiety, the majority classification for each segment of the message sent by the user has to be under the classification of “fear”. Determining if a user has bipolar disease is a little more complicated. In order to determine this condition, the heuristic sees if the message consists of a very similar amount of classified sentiment of “sadness” or “fear” and sentiments of “joy” to see if the message constantly switches between these sentiments. If the message contains a similar amount of segments that are classified as “sad” or “fearful” as segments classified as “joy”, then the patient is diagnosed with bipolar disorder. The chatbot also determines that the patient has no mental illness if the patient does not fall into any of the mental illness categories defined.

Complexity and Effort

Our project used a wide variety of technologies and frameworks to accomplish our goal of creating the chatbot. Our chatbot is segmented into 4 segments, with each segment being built independently until final integrations.

Firstly, the front-end app was built in react-native to facilitate the app’s usability in both android and apple. Further, the frontend app was built in Typescript. The data required for the front-end app was provided by the backend and was accessible through AJAX calls to the backend API endpoints. Second, the backend was built in Python, specifically using the Flask library. Further, the database required to store information on the backend was a PostgreSQL database hosted on a Heroku app. The backend served as the intermediate layer between the frontend and the ML model. Third, the physical illness predictor was a machine learning model built in python, using libraries NumPy, pandas, and scikit-learn. Further, the physical illness prediction model uses a combination of two trained models trained on a Kaggle physical illness dataset, the first model is a random forest classifier, while the second one is a neural network. Further, the model also performs natural language processing to extract relevant information from a user’s message (such as user symptoms) using a combination of a word count and tf-idf vectors. The last segment of the app is the Mental Illness Predictor. The mental illness predictor was also built in python with the use of NumPy, pandas, and scikit-learn libraries. The model uses a Logistic Regression linear classifier to predict illnesses given a list of symptoms. The model used was trained on sentiment analysis datasets of mental illnesses and associated symptoms. Further other sentiment analysis datasets, specifically one performing sentiment analysis on Tweets was used to enhance symptom extraction from a user’s message.

References

- [1] Rosenberg, J. (2020, July 30). Mental health issues on the rise among adolescents, young adults. AJMC. Retrieved February 7, 2022, from <https://www.ajmc.com/view/mental-health-issues-on-the-rise-among-adolescents-young-adults>
- [2] Centers for Disease Control and Prevention. (2021, June 28). About mental health. Centers for Disease Control and Prevention. Retrieved February 5, 2022, from <https://www.cdc.gov/mentalhealth/learn/index.html>
- [3] Mental health by the numbers. NAMI. (2021, March). Retrieved February 5, 2022, from <https://www.nami.org/mhstats>
- [4] Uban, A.-S., Chulvi, B., & Rosso, P. (2021). An emotion and cognitive based analysis of mental health disorders from social media data. Future Generation Computer Systems, 124, 480–494. <https://doi.org/10.1016/j.future.2021.05.032>
- [5] Zunic, A., Corcoran, P., & Spasic, I. (2020, January 28). Sentiment Analysis in health and well-being: Systematic review. JMIR medical informatics. Retrieved February 5, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7013658/>
- [6] Saxena, N. (2020, September 6). Extracting Keyphrases from Text: RAKE and Gensim in Python. Extracting Keyphrases from Text: RAKE and Gensim in Python | by Nikita Saxena | Towards Data Science. Retrieved April 27, 2022, from <https://towardsdatascience.com/extracting-keyphrases-from-text-rake-and-gensim-in-python-eef0fad582f>