CSCE 585: Machine Learning Systems: Project Proposal: Narrative Analysis

September 20, 2024

Contents

Feedback and Responses

Feedback: Ensure our data storage + communication is HIPPA approved **Response**: Making sure the data collection we perform with patients follows all HIPPA guidelines will be one our top priorities

Feedback: Keep in mind the load we are going to be subjecting onto phones/laptops **Response**: While newer phones do have NPUs, and some laptops do as well, it is understandable that some devices may try to run AI locally without an NPU. While we will see how optimized and accessible we can make the ML system, it may be something we have to scale back or limit to devices that contain the proper hardware

Feedback: Look at other datasets (handwriting, facial recognition, video recognition, etc.) **Response**: While handwriting is an interesting idea to consider, the scope of this project is going to be analyzing what the patient is writing rather than how they are writing it. As far as facial recognition and video are concerned, they most certainly would provide invaluable data to the ML system and doctors who use it, but this could cause big privacy concerns, and patients would be less inclined to use the app if it took videos of them in times of weakness.

Feedback: Look at the language used in context **Response**: Everyone has their own unique way of describing how they feel. The goal of our ML system is to identify the variation in language used between different patients. Context is important and our ML system ideally will be able to determine that accurately.

Feedback: Could models be further trained based on region/dialects to create different baselines for different areas **Response**: This is another interesting idea that seems like it may have some promise. However, that it is outside the scope of our current project

Feedback: Could we use tools like apple mental health insights instead of gad-7 and phq-9 (more numerical data and better user interactions) **Response**: While Apple mental health insights may possibly be more effective or popular than GAD-7 and PHQ-9, we believe that sticking to the widely accepted and used methods in healthcare is important as it will make the ML system easier to understand for people already utilizing tools in the healthcare space. On top of that, some doctors may outright deny using a system with a newer measuring metric rather than a well-established one.

Feedback: Keep in mind the scope, build a great project where we just prove the concept, it can always be further developed if we pursue it **Response**: Most certainly! While lots of great ideas have been suggested to us and we continue to discover more each day, we have set a specific goal that we are not going to drastically change.

Project Repository

https://github.com/csce585-mlsystems/Narrative-Analysis/tree/main

Introduction

One of the largest issues in the psychiatric industry is the time it takes to diagnose a patient. Due to the nature of therapy and psychiatric work, provider's need significant amounts of time and experience with any one patient to understand their emotions and provide an accurate diagnosis. By correlating patient's language through daily journals with quantitative surveys, we can provide therapists and psychiatrists with significantly more data on their patients than would've been possible otherwise.

Problem Statement

We want to find a local and cloud-based hybrid system to process health data in a secure, efficient, and private manner. Through ratings of one's emotions through a survey similar to the PHQ-9 and daily journal entries we can correlate textual data to quantitative points and build a history of a patient's language through journaling and use the emotions associated with that language. Hopefully, this would allow providers a better understanding of how their patients are feeling or the problems they face on a day-to-day basis.

Technical Approach

When it comes to tackling this problem, the main approach we plan on utilizing will be a federated learning system with text analysis AI. One of the problems we aim to tackle is that patient data is unique, and training an AI on one patient's data would train the AI ideally for use with that one patient, but trying to use that same AI with other patients could yield drastically incorrect results. This is where federated learning will be effective, federated learning consists of having a "main" AI that is trained on the dataset of other "local" AIs that are only trained on their unique users data. This type of approach will allow an AI to be trained for all patients to be used as a baseline towards any new patients joining the federated learning network, while still preserving the personal trained AIs for each individual patients. Federated learning also helps preserve the privacy of individuals' health data

As a user's AI is trained over time, it will be able to deliver more accurate results for that patient's health data. The AI will analyze patient journals and learn which statementss correspond to which emotions a patient is feeling and how intense those emotions are. Using the data gathered, the AI will be able to generate an actionable list of information about the patient's emotional state for a psychiatrist.

Evaluation

We will evaluate the results by taking the quantitative data gathered by the polls and relating it to the intensity and amount of uses of different emotion-related words to develop a "score" for how strongly an emotion was felt on any given day. By tracking these scores over a period of time for each emotion, we develop actionable insights for providers. For example, things like disjointed spikes in "happy" and "sad" emotions can indicate bipolar disorder, or continuous "sad" emotions with little else can be indicative of major depressive disorder.

Related Work

Facial recognition to determine emotion: Gowda, Ravish, et al. "Artificial Intelligence Based Facial Recognition for Mood Charting among Men on Life Style Modification and It's Correlation with Cortisol." Asian Journal of Psychiatry, vol. 43, June 2019, pp. 101–104, https://doi.org/10.1016/j.ajp.2019.05.017. Accessed 25 Nov. 2019.

Machine Learning for Cognitive Behavioral Analysis: Bhatt, P., Sethi, A., Tasgaonkar, V., Shroff, J., Pendharkar, I., Desai, A., Sinha, P., Deshpande, A., Joshi, G., Rahate, A., Jain, P., Walambe, R., Kotecha, K., & Jain, N. K. (2023, July 31). Machine Learning for Cognitive Behavioral Analysis: Datasets, methods, paradigms, and Research Directions. Brain informatics. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10390406/

References

Bhatt, P., Sethi, A., Tasgaonkar, V., Shroff, J., Pendharkar, I., Desai, A., Sinha, P., Deshpande, A., Joshi, G., Rahate, A., Jain, P., Walambe, R., Kotecha, K., & Jain, N. K. (2023, July 31). Machine Learning for Cognitive Behavioral Analysis: Datasets, methods, paradigms, and Research Directions. Brain informatics. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10390406/

Arqarni, A. (2023). Enhancing Cloud Security and Privacy With Zero-Knowledge Encryption and Vulnerability Assessment in Kubernetes Deployments. Middle Tennessee State University.

Izenman, A. J. (2008). Modern multivariate statistical techniques regression, classification, and Manifold Learning. Springer New York.

Bhatt, P., Sethi, A., Tasgaonkar, V., Shroff, J., Pendharkar, I., Desai, A., Sinha, P., Deshpande, A., Joshi, G., Rahate, A., Jain, P., Walambe, R., Kotecha, K., & Jain, N. K. (2023, July 31). Machine Learning for Cognitive Behavioral Analysis: Datasets, methods, paradigms, and Research Directions. Brain informatics. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10390406/

Arqarni, A. (2023). Enhancing Cloud Security and Privacy With Zero-Knowledge Encryption and Vulnerability Assessment in Kubernetes Deployments. Middle Tennessee State University.

Izenman, A. J. (2008). Modern multivariate statistical techniques regression, classification, and Manifold Learning. Springer New York.

Submission

Please use GitHub tags for indicating your submissions. So, one of your team members submits a `PDF file` that includes, a `title` for your project, a `list of`

members and their `GitHub accounts`, a link to your `repository`, and a specific `GitHub tag` that I should use for locating your submission on GitHub.