CS 510 - Project Final Report

Reddit BDI - Team 21 (TDZ)

Team Members: Tanvi Shah (tanviss2), Darci Peoples (darciap2), Zhaojie Tang (zhaojie2)
Project Track: Application Track

1 Significance

According to the World Health Organization (WHO), depression is a common illness worldwide, with more than 264 million people affected [2]. The severity of depression can vary vastly from person to person with the milder symptoms often being extremely hard to detect. It is extremely important to detect depression and mental health issues in time, as these can lead to various physical and mental conditions in the long run. Some of the symptoms include lower energy levels, lower levels of interest in activities one would otherwise find interesting, loss of appetite or overeating, changes in sleep cycle and other changes in daily behavior and communication. If detected early, medicines and therapy can get rid of depression and largely improve mental health conditions.

In a time where social media and online communication has become extremely popular, social media posts pose as an invaluable source of information. We believe that one could use information from a person's social media posts and predict if that person is likely to have depression or not. Using posts can potentially result in an early detection of symptoms of depression in people and we think that a risk assessment tool for depression would be a helpful resource for people, both users as well as researchers and/or therapists.

"One of the major problems with depression is that most people don't recognize that that's what they are experiencing," said Dr. Gary Kaplan, founder of The Kaplan Center for Integrative Medicine, located in McLean, Va, in a statement. "Once we know that someone is depressed, we can take an integrative approach to set them on the path to wellness. But they need to take that first step of coming into our office."[1] Our application offers a reference for people about their own depression status. Those who found out their result to be severe could contact a physician for an evaluation in time.

2 Users

We hypothesize various different kinds of individuals that could be potential users of our application. The first, obvious user type are individuals who wish to assess themselves to check if they might be suffering from depression or not. These users can feed in their Reddit usernames on our Application's home page. Our system will then retrieve several of the user's posts, run it through our backend model, and auto-feed in the predicted answers to the BDI (Becks' Depression Inventory) questionnaire [3]. The cumulative results of these answers are also

[1]Dr. Kaplan's quote: hcplive.com/view/patients-unaware-physical-and-emotional-symptoms-related-to-depression

converted into 4 categories by our system: Minimal depression, Mild depression, Moderate depression and Severe depression.

Apart from this, we think our system can also be of potential use to researchers and other health-care practitioners who wish to analyze statistics regarding depression and its severity among humans. Therapists can use this platform to assess and gain insights about their patients. Our platform also works for people who do not have reddit accounts or who do not want to use their reddit history to predict answers to the questionnaire. We provide the BDI form on our Web Application, where users can manually fill in the answers to the questions and see the results. They can also share their results with us, which can help improve our system further.

3 Novelty

Specific to the BDI questionnaire, we mostly just found websites that provided a non-interactive version of the questionnaire in the form of downloadable pdf files, where users would have to download the file, fill in the questionnaire and calculate the results based on the provided instructions. This requires too much effort by the user and they might get discouraged or disinterested to follow through.

There were a few websites we found that did have interactive forms using the BDI questionnaire, and one could potentially fill out the answers to each question, and check their results. However, we feel that when a user is asked to think about and answer a question, their thought process and frame of mind shift, and this could cause their answers to be biased.

There has also been research into the early detection of major depressive disorder on Twitter and detecting depression through Facebook statuses. There has been less research done in mental health checks through Reddit.

Our system tries to analyze a user's Reddit posts and interaction history to auto-assess them for risks of depression. We have not found any other publicly-available websites or systems that claim to do the same thing we propose to do in our system, and in this regard, we believe that our system would be different than already existing systems.

4 Evaluation of the Model

To predict answers to the BDI Questionnaire, we make use of a BERT based model using Hugging Face Transformers. We built our model using TensorFlow Keras. The input to our model is a collection of text from a user's most recent Reddit posts. The text includes the title of the post, and the actual content of the post. The problem under consideration was essentially a Multi-Class Multi-Label text classification problem. There are 21 classes in total, corresponding to the 21 questions in the questionnaire. Each of the 21 questions can have 4 answers, namely 0, 1, 2 and 3, except questions 16 and 18 which have 7 possible answers: 0, 1a, 1b, 2a, 2b, 3a and 3b. Thus for a given input from a single user to the model, we want to predict a label for each one of the 21 classes (questions).

As part of our project, we expanded on the model from our experimental lab (Task 2.3). From the CLEF eRisk 2021 Workshop[2], we had data for about 90 users. We split this data into a training set and a test set with 75 users for training and 15 users for testing. We used a pre-trained BERT model, and instead of the one dense layer at the end of the model, which is used if there is exactly one class we are working with, we had to add 21 different dense layers – one layer for the prediction of each class/question. We used an Adam Optimizer and the Categorical Cross Entropy loss to calculate the loss metric.

We also used the Bert model's inbuilt tokenizer to tokenize the inputs. The maximum input size for a BERT tokenizer is 512. So any input text that is longer than 512 is just truncated. Since the collection of all posts by a given user is bound to be more than 512 words in length, we wanted to find a way to not lose a bunch of our information. So, instead of passing the cumulative text from all the posts of a given user as a single input, we pass individual posts as the input. Thus, our model predicts separate answers to the 21 questions for each post of a given user. To get a single result for the user, we pick the average answer from all the posts for a given question.

We evaluated our model based on the following different metrics:

1. Average Hit Rate (AHR): This can be calculated using the number of overlapping correct answers as predicted by the model with the ground truth answers
2. Difference between Overall Depression Levels (DODL) and Average DODL: Overall depression level is calculated as the sum of all the answers of the questionnaire (ranging between 0 and 63). DODL is the difference in depression levels as predicted by the model and the ground truth.
3. Depression Category Hit Rate (DCHR): This metric is calculated using the number of correct predictions for the different categories of depression: minimal, mild, moderate and severe.

We submitted our model to the CLEF eRisk 2021 competition under Task 3 and we received evaluation results for our model on a hidden test set. The T3 results are in the following table[3]: https://drive.google.com/file/d/1plHQQh3UnKpuNqrUYDReI9XxqtqE5G9p/view

Here a summary of our results, specifically:

	AHR	**ACR**	**ADODL**	**DCHR**
Our metric	35.13%	67.76%	75.81%	22.50%
Our rank	#3	#32	#17	#18
erisk2021 best	38.78%	73.17%	83.59%	41.25%

From the above results, we can conclude that our model performed pretty well in the AHR metric (among the top scoring compared to the other submitted models). We can also see that there is some room for improvement, especially in the DCHR metric. We try to improve the

[2] eRisk 2021: https://erisk.irlab.org
[3] You may view the entire erisk2021 Evaluations Results pdf here: https://drive.google.com/file/d/16Lx2yro8ovWBB7XQ0OxFgk0zCBFwlXrv/view

current model through customized data preprocessing and the usage scheduled cyclic learning rate, but the result for the moderate changes insignificantly. There is a possible future work and improvement for our current system, such as incorporating reliable feedback to update the model.

5 Evaluation of the UX

Our application is a webapp to let users estimate their depression severity using their Reddit comment & post history. With our application, users can:

1. Fetch their 200 most recent Reddit posts and comments
2. Auto-complete the BDI for a low-effort estimate of depression severity
3. Manually fill responses to the BDI in order to gauge their depression severity
4. View depression level equivalent to their responses
5. Share their responses with us to potentially improve the system
6. Browse a catalog of mental health resources[4]

The application was made using React, Bootstrap, Flask, Python, TensorFlow, PushShift.io, and GCP services. React and Bootstrap helped us make a nice user interface. We further improve the UX by providing alerts to let users know when a request is being served, has failed, or was successful.

Readers can download a flow-chart and **view screenshots of our user story** here[5]: https://drive.google.com/drive/folders/1w893Y-DFnAvkJTA1KkCZjIC34AiQzxbg

5.1 Live Application

You can find our live application here: https://reddit-bdi.uc.r.appspot.com

Alternatively, you can download & run our source code, following the README.md: https://drive.google.com/drive/folders/1Fcv6YOMdq1UHrMYMKQ1jqyALcXXl-oK4

There are a few limitations to remember when using the system:

1. When fetching posts and comments from Reddit, allow up to 10 seconds.
2. When the server is predicting responses, allow up to 120 seconds. It takes a while to load the model, tokenize the input, and finally predict questionnaire responses.
3. Don't spam the system. The live app can only accept a few concurrent requests. If you see a red error alert, give the system a few minutes to recover before trying again.

5.2 Example Case Studies

Currently, a user can search for any Reddit user's posts, estimate their depression level, and submit responses. This is a big privacy & security concern that we would need to address before ever opening this to real users. For assessment purposes, you can test with these users. The users we list here are either public personalities that posted an "ask me anything" on r/IAmA or users

[4] Resources pulled from https://csmhcillinois.com
[5] Project progress video shows the user story too: https://mediaspace.illinois.edu/media/t/1_n7yt73n9

found through the r/SuicideWatch and r/Depression subreddits whose names were redacted for privacy. These examples show some combinations of under/over/good estimates for people who are or are not actually depressed.

- *<username removed for privacy>*: An active user of r/SuicideWatch and r/BPD. We've asked Reddit to reach out to them due to concern for their wellbeing. They have the highest depression score we've found (moderate), but we believe they may very well have severe depression. This exemplifies our low metrics for DCHR.
- *ElonMuskOfficial*: Elon Musk's official account. He has few submissions that are all technical. He receives one of the lowest depression scores we've seen (minimal). We think this is accurate.
- *PresidentObama*: Former President Barack Obama's account used for an AMA. He has few submissions, all dealing with politics, directed at voters. Our model estimates a mild case of depression. We think this may be an overestimation possibly due to the subject or few number of posts.
- *SethEverman*: A medium-sized YouTuber who has talked about having a "light" case of depression. Our application predicts mild depression--accurately so.

6 Task Division

Darci Peoples used PushShift.io to fetch a given Reddit users' posts and comments. She also built the webapp and entire user experience (UX) for our system using React, Flask, and Google App Engine.

Tanvi Shah researched on different methods and existing tools for NLP and built the basic BERT model that we use for predicting the answers to the BDI questionnaire. She also figured out how to store our trained model on the cloud and call it using our deployed Web Application.

Zhaojie Tang worked on various methods to improve model performance, using techniques such as appropriate pre-processing of data and the usage of scheduled and cyclic learning rates for the Adam optimizer of our BERT model. He also collected evaluation metrics for our model.

References

[1] C. (2020, February 27). Training your Neural Network with Cyclical Learning Rates –. MachineCurve. https://www.machinecurve.com/index.php/2020/02/25/training-your-neural-network-with-cyclical-learning-rates/

[2] "Depression Overview." *World Health Organization*, World Health Organization, www.who.int/health-topics/depression#tab=tab_1.

[3] Beck, A. T., Steer, R. A., & Brown, G. (1996). *Beck Depression Inventory–II* [Database record]. APA PsycTests. https://doi.org/10.1037/t00742-000