DREU 2023 Final Report

Ty Mercer

Project 1.....**2**

History of Project..... 2

Where I Contributed.....2

Future Research..... 3

Project 2.....**3**

Project 3.....**4**

History Of Project.....4

Where I Contributed.....4

Future Research..... 5

References..... **6**

Project 1

Machine Learning to Determine if Someone is Suicidal, to be Used in Suicide Prevention

History of Project

In January 2023, three student researchers began working on the project. Their objective was to create a language model that could predict a social media user's risk of suicide based on their posts.

The researchers quickly found a publicly available pre-trained model called MentalBERT [1] that could be used for this task. MentalBERT is based on the BERT model created by researchers at Google for natural language processing [2]. MentalRoberta

Researchers also found a dataset from the University of Maryland available for the training of the model. This dataset includes Reddit users and their posts, along with expert or crowd-sourced evaluations of 4 levels of suicide risk.

By June 2023, the research itself had finished, and researchers were readying for submission into a high-tier machine learning conference. The researchers chose to use a variant of MentalBERT called MentalRoBERTa because of higher accuracy, and used TF-IDF scaling on each token to improve accuracy. With these improvements, researchers were able to achieve an accuracy of 63.05%. The paper was submitted to a top-tier machine learning conference in July.

Where I Contributed

I got together the related work for the academic paper we submitted to a top-tier ML conference. This included going through pages of Google Scholar and retrieving only articles that could connect to our research. This also included the easier task of using the Scholarcy tool to generate summaries from the PDFs of each article. In gathering the related work, I showed other academics reading the paper that this project's researchers are cognizant of present research.

I write the documentation for each script and resource in the project. This includes determining the input and output of the script to understand how it functions in the larger project. I have written most of it, and I have made a commitment to my mentor to finish it by the end of

August. Because of this documentation, it will be easier for future researchers in the project to contribute.

I have done preliminary work to act as a project manager and paper writer for the next research paper related to this project. This includes meeting with a researcher to understand what is required next, and communicating with several researchers for important meetings required to move forward with the paper. This paper will be further described in the next section. I have made a commitment with my mentor towards its completion this Fall. My management and writing of this paper will allow a researcher to see through with his research ideas before he has to move away from the project.

Future Research

One researcher in the project has an idea for another paper that I mentioned earlier. In this paper, we will pair MentalRoBERTa (the model making predictions) with a language generation model similar to ChatGPT in how it behaves. The paper will focus on the prompt engineering aspect of letting the language generation model know the prediction from MentalRoBERTa before looking at the posts, then having the language model give reasons for why the user and their posts show the predicted risk of suicide. I will act as the project manager and one of the main writers for the paper this Fall. In doing so, the commitment to a certain amount of hours from DREU will be fulfilled.

Project 2

Machine Learning Training To Bolster My Knowledge

When I was offered the DREU program opportunity by my mentor Dr. Nasheen Nur, we were initially planning for me to contribute to Project 1 by assisting in scripts that directly involve machine learning. However, by the time I began working on the project, most of this work was already completed or delegated to other researchers who could more efficiently develop these scripts. With this in mind, my mentor and I decided that I would spend time in machine learning training to better develop my knowledge in the area.

The training I have worked on has given me insight into using the most prevalent python libraries or tools necessary for machine learning. It has also given me insight into the conceptual

basics for machine learning. I learned about the Weka tool and how to tune algorithm parameters, rescale parameters so they fall within a desired range, and perform feature selection. I learned about the interconnectedness of variance and bias, and the pros and cons of parametric and non-parametric models. I learned about tools to debug and profile Python code. I learned how to visualize data in plots and charts with matplotlib or seaborn. I learned how to set up k-fold cross validation and models with scikit-learn, as well as tune parameters.

Project 3

Assisting With The Development of a Math Education App.

History Of Project

During the Spring 2023 Semester, Victor Tumbiolo started working on math animations to clarify math concepts taught in high school. The animations clarified math concepts such as the Maclaurin series, the properties of a circle, and the unit circle. Victor is another student in the DREU Program, and we have the same research mentor, Dr. Nasheen Nur. His primary research objective for the DREU Program became creating new animations in Java and compiling them into a web application. With guidance from our mentor, he envisioned creating an academic paper around this application. As of now, he is improving his animations after feedback, and is getting code set up for the web application.

Where I Contributed

I wrote feedback after reviewing his animations. I included ideas like how to make it more clear what each math term refers to, such as how to make it clear what a sector is in the properties of a circle animation. I made note of minor visual glitches, such as a part of the y-axis not showing up after drawing the area between two curves. Finally, I provided suggestions to use color to differentiate between math ideas, such as coloring the area between two curves a different color than the area above one curve and the area below another.

I wrote a pre-survey that asks about someone’s high school math education up to this point and where they believe their math education could be improved. This was completed so

that the web application can be part of a larger paper that is compliant with academic standards of research. See the “Future Research” section immediately below for more.

Future Research

Victor plans to have the web application, pre-survey, and post-survey sent out to both high schoolers and college students this Fall. In doing so, he will fulfill academic standards for research into better math educational tools.

References

[1] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[2] Ji, Shaoxiong, Tianlin Zhang, Luna Ansari, Jie Fu, Prayag Tiwari, and Erik Cambria. "MentalBERTt: Publicly available pretrained language models for mental healthcare." arXiv preprint arXiv:2110.15621 (2021).

[3] H.-C. Shing, S. Nair, A. Zirikly, M. Friedenberg, H. Daumé III, and P. Resnik, “Expert, crowdsourced, and machine assessment of suicide risk via online postings,” in Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, 2018, pp. 25–36.

[4] A. Zirikly, P. Resnik, Ö. Uzuner, and K. Hollingshead, “CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts,” in Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology, June 2019.