Human-AI Collaboration for Mental Health

Geovanni Cuevas, Yousef Hasan, Abm. Adnan Azmee, & Md Abdullah Al Hafiz Khan

College of Computing and Software Engineering

KENNESAW STATE UNIVERSITY

Introduction

Mental health has become a pressing global concern, influencing cognition, emotions, and social interactions. With social media platforms like Reddit serving as vital arenas for mental health discussions, understanding these conversations presents challenges due to their nuanced and varied nature.

Our project, Human-AI Teaming for Mental Health, proposes a unique approach to address this complexity. By combining human expertise with artificial intelligence, we aim to annotate Reddit posts based on factors contributing to mental health conditions. Utilizing advanced deep neural network models and curated datasets, we seek to effectively analyze and interpret these discussions.

By integrating human insights into AI-driven analyses, we anticipate improved identification of underlying mental health issues. Our research holds significant implications for healthcare professionals, researchers, and policymakers, offering actionable insights to support mental well-being in the digital sphere.

Method

- Data Acquisition: Reddit data pertaining to mental health discussions was collected in JSON format. The dataset encompassed a diverse array of posts, each encapsulating insights into users' perceptions and experiences related to mental health.
- Data Preprocessing: The acquired JSON files were preprocessed using Python's Pandas library. This involved cleaning the data, extracting relevant features, and structuring it into a format conducive for training the neural network.
- Neural Network Architecture: A neural network was constructed utilizing the Keras framework. This architecture was designed to process the structured data extracted from Reddit posts. The network comprised multiple layers, including input, hidden, and output layers, configured to effectively capture the intricate patterns inherent in the dataset.
- Training Process: The constructed neural network was trained using the preprocessed Reddit data. Initially, the network underwent training using standard training techniques. Subsequently, Long Short-Term Memory (LSTM) units were incorporated into the network to enhance its ability to capture temporal dependencies within the data.
- Training Loss Curve: During the training process, the loss function—indicative of the disparity between predicted and actual values—was monitored and recorded over successive epochs. A training loss curve was plotted to visualize the convergence of the network's performance over time, facilitating the assessment of model training efficacy and convergence stability.
- Fine-Tuning with LSTM: Following the initial training phase, the network was further fine-tuned using LSTM units. This iterative process enabled the network to refine its understanding of temporal sequences inherent in Reddit discussions, thereby enhancing its predictive capabilities and generalization performance.

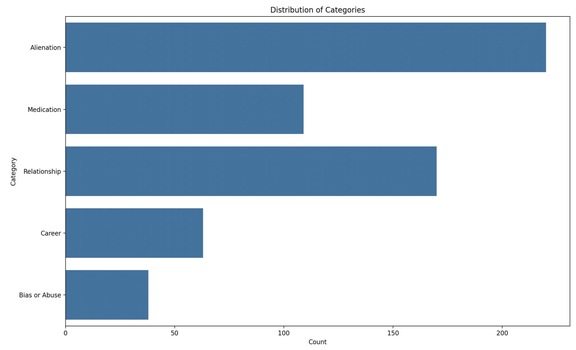
Results

Our model achieves an average accuracy of 43%, with Career standing out at 48% and Alienation at 45%. To improve accuracy, addressing data imbalance and ensuring data quality are essential. Further enhancements may involve refining model complexity, optimizing hyperparameters, and incorporating domain-specific features. Ensemble methods, regularization, and transfer learning could also boost performance. However, perfection may remain elusive due to the complexity and subjectivity of mental health topics. Nonetheless, by implementing these strategies, we aim to enhance our model's accuracy in classifying mental health-related text, providing valuable insights into these crucial areas for mental health clinicians.

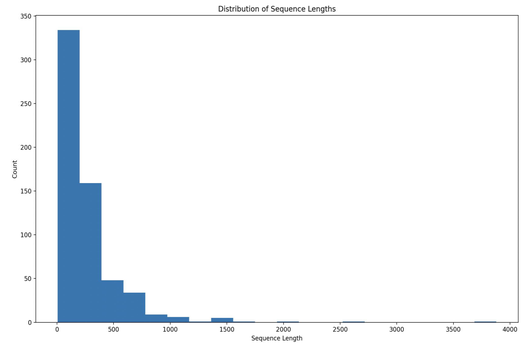
Conclusion

Mental health, a global concern, is intricately discussed on platforms like Reddit. Our project, Human-AI Teaming for Mental Health, merges human expertise with AI to annotate mental health-related Reddit posts. Using advanced neural network models and curated datasets, we aim to effectively analyze these discussions. By integrating human insights into AI analyses, we anticipate improved identification of mental health issues. Our research has significant implications for healthcare, AI research, and policy. It offers actionable insights to support mental well-being in the digital sphere, aiding in targeted interventions and policies. Ultimately, our approach empowers people with valuable information to address mental health challenges in the digital age.

Distribution of Data



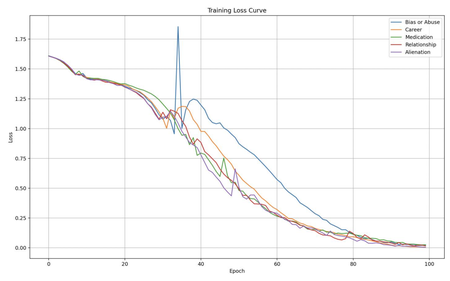
Distribution of Sequence Lengths



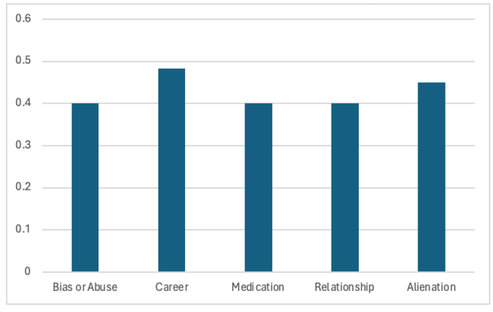
Output in Program

```
Training model for Bias or Abuse...
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 24ms/step - accuracy: 0.4229 - loss: 3.6121
Test Accuracy for Bias or Abuse: 0.4000000059604645
Training model for Career...
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 29ms/step - accuracy: 0.4681 - loss: 3.4893
Test Accuracy for Career: 0.4833333194255829
Training model for Medication...
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 34ms/step - accuracy: 0.4333 - loss: 3.3348
Test Accuracy for Medication: 0.4000000059604645
Training model for Relationship...
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 38ms/step - accuracy: 0.4125 - loss: 3.6044
Test Accuracy for Relationship: 0.4000000059604645
Training model for Alienation...
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 26ms/step - accuracy: 0.4875 - loss: 2.7645
Test Accuracy for Alienation: 0.44999998807907104

Average Accuracy: 0.42666666507720946

Process finished with exit code 0
```

Training Loss vs Epochs Curve



Class-Wise Accuracy



References

Tang, Yujian. “Long Short-Term Memory (LSTM) in Keras.” *PythonAlgos*, 23 Jan. 2023, pythonalgos.com/long-short-term-memory-lstm-in-keras/.

bhaveshkumar2806. “Complete EDA and Visualization of Text Data.” *Kaggle*, Kaggle, 5 June 2021, www.kaggle.com/code/bhaveshkumar2806/complete-eda-and-visualization-of-text-data.

Pandas Read CSV, www.w3schools.com/python/pandas/pandas_csv.asp. Accessed 9 Apr. 2024.

GfG. “Read JSON File Using Python.” *GeeksforGeeks*, GeeksforGeeks, 27 July 2023, www.geeksforgeeks.org/read-json-file-using-python/.

Acknowledgments

Kennesaw State University, College of Computing and Software Engineering

First Year Scholars Program

Vertically-Integrated Project (VIP) Program