UDAN: Unobtrusive Sentiment Alert Using Natural Language Processing

Sampathkumar, Anagha (School: National Public School, Indiranagar)

Aim To develop a novel Natural Language Processing tool to detect depression early based on users' social media activity. Methods and materials A corpus of 50,000 text data samples from posts and comments from r/Anxiety Depression and r/mildly interesting were acquired from Reddit. We also maintained a repository of 100 unlabeled samples with ambiguous sentiment to further test the model's accuracy. We preprocessed the data by standardizing it into lowercase and removing punctuations and emoticons. Next we tokenized the data by Sub-Word Tokenization using RegEx method and Python Split Function, subsequently loading it into a custom dataset. We used a BERT-based model as it is pre-trained and could represent the data bidirectionally, allowing interpretation of a sentence's context rather than interpreting it word-by-word. Finally, we trained and tested the model with our dataset with a learning rate of 0.00002, batch size of 32 and 5 epochs. We evaluated our model on its accuracy, loss, precision, recall and F1-score. Results Based on rigorous testing, UDAN displays high accuracy of 92.88%, loss of 0.26, precision of 0.87, recall of 0.88 and F1-score of 0.88. UDAN's dependability stems from its analysis of depressed sentiment in the overall context of the sentence rather than individual words. Further, UDAN displayed a high accuracy of 91% on the unlabeled data with ambiguous sentiment. Conclusion UDAN is quantitative, unbiased, economical and accurate, enabling individuals who suspect they are depressed to get an early detection, creating the first step in a multi-phase screening process. With UDAN, AI can complement conventional diagnostics so patients receive the right action at the right time.

Awards Won:
Fourth Award of $500