CONCLUSION

In this paper, we designed and evaluated a novel approach to monitor the mental health of a user on Twitter. Building off existing research, we worked to translate and quantify suicide warning signs in an online context (user-centric and post-centric behavioral features). In particular, we focused on detecting distress-related and suicide-related content and developed two approaches to score a tweet: an NLP-based approach and a more traditional machine learning text classifier.

To detect changes in emotional well-being, we considered a Twitter user’s activity as a stream of observations and applied a martingale framework to detect change points within that stream. Our experiments show that our NLP text-scoring approach successfully separates out tweets exhibiting distress-related content and acts as a powerful input into the martingale framework. While the martingale values “react” to changes in online speech, the change point detection method needs improvement. We were able to detect the true change point for one validation case, but the approach needs to be more robust with respect to parameter setting and positive changes in speech.

For future research, we plan to further explore the impact of martingale parameters on the change detection effectiveness. We also hope to expand the approach to include image processing and other social media outlets in order to assess the effectiveness in other settings. Another interesting perspective is to consider more fine-grained emotion classes such as anger, sadness, fear, etc., instead of considering four levels of distress. However, overall, we believe our initial work presents an innovative approach to detecting suicide-related content in a text stream setting.