Report: AI for Good and Devil's Advocate Analysis
Assessment 4

Artificial Intelligence
COSC550

Submitted By
Bikash Neupane
220245756

Submitted To
Raymond Chiong
Unit Coordinator,
UNE

Table of Contents

Report: AI for Good and Devil's Advocate Analysis

In this report, I have critically examined two research papers that propose AI-driven solutions for detecting depression using social media texts. Both papers outline methodologies for depression detection through machine learning (ML) approaches, emphasizing the potential for positive societal impact. Additionally, I explored the potential risks and negative consequences of these AI technologies, adopting the role of a devil's advocate to highlight concerns around privacy, ethics, and future deployments. This dual analysis aims to enhance my understanding of the ethical considerations surrounding AI's role in mental health.

Paper 1: A textual-based featuring approach for depression detection using machine learning classifiers and social media texts

Objectives and Research Questions

The first paper, authored by Raymond Chiong, Gregorius Satia Budhi, Sandeep Dhakal and Fabian Chiong, investigates whether machine learning techniques can detect depression in social media users by analysing their posts, even when these posts do not explicitly mention depression-related keywords. This paper addresses a critical issue in mental health: undiagnosed depression due to lack of awareness or social stigma. The primary objective is to determine if machine learning models, trained on textual features, can identify early signs of depression without relying on specific keywords like "depression" or "diagnosis" (A textual-based featuring approach for depression detection using machine learning classifiers and social media texts, pp. 1, 2).

Methodology and Techniques

The methodology involves training and testing several machines learning models, including both single classifiers (Logistic Regression, Support Vector Machines) and ensemble methods (Random Forest, Gradient Boosting), using two publicly available Twitter datasets. The labels "Depression" and "Non-Depression" were applied to these datasets. The datasets were subjected to several text pre-processing techniques, including lemmatization, tokenization, and stop word removal. Furthermore, sentiment analysis and n-grams were employed to improve the models' predictive power (A textual-based featuring approach for depression detection using machine learning classifiers and social media texts, pp. 2, 3, 4, 5).

To evaluate the generalizability of these models, they were also tested on non-Twitter depression-class datasets sourced from Facebook, Reddit, and personal diaries. The authors explored whether their model could detect depression even in posts where users did not explicitly express their mental state (A textual-based featuring approach for depression detection using machine learning classifiers and social media texts, pp. 5, 6, 7, 8, 9).

Key Findings and Results

The study demonstrated that machine learning classifiers could successfully detect signs of depression in social media users, even when depression-specific keywords were absent. However, the results showed that overfitting occurred when models trained on Twitter datasets were applied to non-Twitter data. To mitigate this, the authors removed keyword dependencies and retrained the models, achieving improved performance on unseen data. Ensemble models, particularly Random Forest and Gradient Boosting, showed the highest detection accuracy, with values exceeding 95% on the test datasets (A textual-based featuring approach for depression detection using machine learning classifiers and social media texts, pp. 8, 9, 10).

Intended Positive Impact on Society (AI for Good)

This research could have a major effect on mental health treatments since it identifies individuals who are at risk of depression based on their usage of social media. By using this method, mental health professionals may be able to reach out to individuals who may not be aware that they require care. Depression is a condition that is commonly underdiagnosed. Early detection could reduce the risk of severe mental health outcomes, such as suicide, especially among populations that express their emotions more freely online.

Paper 2: Combining Sentiment Lexicons and Content-Based Features for Depression Detection

Objectives and Research Questions

The second paper, authored by Raymond Chiong, Gregorius Satia Budhi and Sandeep Dhakal, proposes a hybrid depression detection method by combining sentiment lexicons with content-based features. The research aims to explore how this combination can enhance the accuracy of detecting depressive tendencies in social media posts. Specifically, the authors sought to identify whether sentiment lexicons (e.g., SentiWordNet and SenticNet) combined with textual characteristics (e.g., part-of-speech tagging, sentence structure) could outperform traditional feature extraction methods in depression detection (Combining Sentiment Lexicons and Content-Based Features for Depression Detection, pp. 99, 100).

Methodology and Techniques

In this study, the authors introduced 90 unique features derived from sentiment lexicons and content-based attributes of Twitter posts. Sentiment features were extracted from lexicons like SentiWordNet and SenticNet, while content-based features included metrics such as sentence complexity, the presence of negations, and linguistic characteristics. Machine learning classifiers such as Logistic Regression, Decision Trees, and Gradient Boosting were then trained using these features. The models were evaluated using two publicly available datasets, also derived from Twitter posts, labelled as either “Depression” or “Non-Depression” (Combining Sentiment Lexicons and Content-Based Features for Depression Detection, pp. 100, 101, 102).

To make sure the models were strong, a 10-fold cross-validation procedure was applied. The classifiers were assessed using performance criteria such F1 score, recall, accuracy, and precision. The relative success of sentiment lexicon-based features and traditional textual characteristics in predicting depression was examined (Combining Sentiment Lexicons and Content-Based Features for Depression Detection, pp. 101, 102).

Key Findings and Results

The study concluded that combining sentiment lexicon-based features with content-based features significantly enhanced the accuracy of depression detection. Gradient Boosting achieved the best results, with accuracy exceeding 98%. The authors noted that content-based features alone were not sufficient, but when combined with sentiment lexicons, they improved detection performance across all classifiers. The hybrid approach also mitigated the problem of class imbalance present in the datasets, particularly in Eye's dataset, where "Non-Depression" samples heavily outnumbered "Depression" samples (Combining Sentiment Lexicons and Content-Based Features for Depression Detection, pp. 102, 103, 104).

Intended Positive Impact on Society (AI for Good)

The hybrid method proposed in this paper provides a more nuanced approach to detecting depression. By incorporating both sentiment and content-based features, the model can capture subtle cues of depression that might be overlooked in traditional analyses. This could result in more accurate identification of at-risk individuals and earlier intervention. Additionally, because the model is not dependent on the explicit use of depression-related keywords, it could be deployed across a broader range of social media platforms and applied to users who might be less vocal about their mental health struggles.

Devil's Advocate Analysis

Advocacy in Favor of the Proposed AI Methods

Both papers present promising AI methodologies for depression detection with the potential to improve mental health outcomes on a global scale. The capacity to identify depression automatically from social media posts is a significant development in the medical field:

Early Detection and Intervention: Even if a person refuses to get medical attention, AI systems are capable of warning those who are at risk for depression. This may be especially helpful for people that may not be aware of their own mental health issues or in areas with few mental health resources.

Scalability and Efficiency: These AI-powered methods can be used widely, possibly keeping an eye on millions of social media users. This would make mental health care easily available and accessible by enabling healthcare systems to quickly and successfully detect and help those in need.

Personalized Mental Health Care: These technologies have the ability to enable more customized treatment programs and give mental health professionals an understanding of patient behaviours that they might not otherwise discover through examining individual social media use.

Devil's Advocate: Potential Negative Consequences

Although the suggested AI solutions have a lot of potential applications, they also come with major risks and ethical issues:

Privacy Invasion: It may be a serious breach of someone's privacy to monitor social media posts for symptoms of depression without permission. It's possible that users aren't aware that their online behaviour is being examined for signs of mental illness, which could be against informed consent ethics.

Algorithmic Bias: These AI systems run the risk of maintaining or even increasing current presumptions. For example, individuals from minorities or those who speak in separate languages may not be well-represented in the training datasets, which could result in incorrect diagnoses or exclusions from detection.

Over-Reliance on AI: Mental health professionals run the risk of having too much trust on AI-based technologies, which could decrease the importance of human judgment and humanity in providing mental health care. AI can help with diagnosis, but it shouldn't take the place of physically conversations with patients for additional understanding.

Data Misuse: Another issue is the possible misuse of data for profit. The potential for social media platforms to use depression detection algorithms for data mining or targeted advertising raises severe ethical concerns regarding the privatization of mental health.

Conclusion

These studies provide significant advantages for society in terms of prevention and treatment using modern AI approaches for social media text-based depression detection. But there are also serious dangers related to algorithmic bias, ethics, and privacy. To minimize these dangers:

- Before implementing such systems on a large scale, user permission procedures and strong privacy safeguards need to be in place.
- Artificial intelligence models need to be open and constantly evaluated for assumptions and errors.
- AI-driven healthcare solutions should always be supported by human monitoring to ensure that technology improves, not replaces, personalized mental health care.

By addressing these challenges, AI systems can effectively improve mental health performance while minimizing potential risks by handling these challenges.

References

Chiong, R., Budhi, G. S., & Dhakal, S. (2021). Combining Sentiment Lexicons and Content-Based Features for Depression Detection. *Affective Computing and Sentiment Analysis*, 99, 100.

Chiong, R., Budhi, G. S., & Dhakal, S. (2021). Combining Sentiment Lexicons and Content-Based Features for Depression Detection. *Affective Computing and Sentiment Analysis*, 100, 101, 102.

Chiong, R., Budhi, G. S., & Dhakal, S. (2021). Combining Sentiment Lexicons and Content-Based Features for Depression Detection. *Affective Computing and Sentiment Analysis*, 101, 102.

Chiong, R., Budhi, G. S., & Dhakal, S. (2021). Combining Sentiment Lexicons and Content-Based Features for Depression Detection. *Affective Computing and Sentiment Analysis*, 102, 103, 104.

Chiong, R., Budhi, G. S., Dhakal, S., & Chiong, F. (2021). A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. *Computers in Biology and Medicine*, 1, 2.

Chiong, R., Budhi, G. S., Dhakal, S., & Chiong, F. (2021). A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. *Computers in Biology and Medicine*, 2, 3, 4, 5.

Chiong, R., Budhi, G. S., Dhakal, S., & Chiong, F. (2021). A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. *Computers in Biology and Medicine*, 5, 6, 7, 8, 9.

Chiong, R., Budhi, G. S., Dhakal, S., & Chiong, F. (2021). A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. *Computers in Biology and Medicine*, 8, 9, 10.