International Journal of Engineering Research and Science & Technology

IJERST

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(3) (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

FORECASTING NATIONAL-LEVEL SELF-HARM TRENDS WITH SOCIAL NETWORKS

NAGESWARA RAO KOLUSU
nageswarmca07@gmail.com
24NH1D5811
VALLURI SUDHAKAR
vallr.sudhakar@gmail.com
ASSOCIATE PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
V.K.R &V.N.B Engineering College

ABSTRACT

"Forecasting National-Level Self-Harm Trends with Social Networks" presents an intelligent data-driven framework that analyzes large-scale social network activity to identify patterns associated with mental health concerns and predict national-level self-harm trends. The study utilizes machine learning, natural language processing, and sentiment analysis techniques to examine publicly available social media content, behavioral indicators, and temporal trends while maintaining user privacy and ethical standards. By integrating historical health statistics with online social interaction data, the proposed system aims to provide early warning signals for mental health authorities and policymakers to support timely intervention strategies and resource allocation. Experimental results demonstrate that social network signals can significantly improve forecasting accuracy compared to traditional statistical methods, enabling proactive public health planning and enhancing mental health awareness at a national scale.

Received: 20-03-2026 Accepted: 28-04-2026 Published: 04-06-2026

INTRODUCTION

Self-harm and mental health disorders have become significant public health concerns worldwide, affecting individuals across different age groups and societies. Traditional methods of monitoring self-harm trends mainly rely on hospital reports, surveys, and clinical records, which often suffer from delays in data collection and limited real-time insights. With the rapid growth of social networking platforms, people increasingly express their emotions, stress, anxiety, and psychological conditions online, creating valuable digital footprints that can help identify mental health patterns at an early stage.

The emergence of artificial intelligence, machine learning, and natural language processing has enabled researchers to analyze large-scale social media data for detecting behavioral and emotional indicators related to self-harm tendencies. Social networks such as online forums, blogs, and microblogging platforms provide real-time information that can reveal public sentiment, emotional distress, and changes in mental health trends at a national level. By studying these online interactions, researchers can forecast potential increases in self-harm incidents and support preventive healthcare strategies.

The proposed system focuses on forecasting national-level self-harm trends using social network data combined with advanced predictive analytics techniques. The framework collects and processes publicly available social media content, extracts emotional and linguistic patterns, and applies machine learning models to identify correlations between online behavior and self-harm statistics. This approach helps healthcare organizations and policymakers gain timely insights for planning awareness campaigns, allocating mental health resources, and implementing early intervention programs. The system aims to improve forecasting accuracy, enable proactive decision-making, and contribute to the development of data-driven mental health support systems while maintaining ethical standards and user privacy.

LITERATURE SURVEY

1. "Detecting Depression and Mental Illness on Social Media"

Author: Munmun De Choudhury et al.

Description:

This research focused on analyzing social media activities to identify signs of depression and mental health disorders. The authors used linguistic analysis, posting behavior, and emotional patterns from online platforms to predict psychological conditions. The study demonstrated that social media data can be effectively utilized for early mental health detection and public health monitoring.

2. "Social Media as a Tool for Monitoring Mental Health Trends"

Author: Glen Coppersmith et al.

Description:

The study explored how machine learning techniques can analyze large-scale social network data to detect mental health conditions such as anxiety, stress, and suicidal thoughts. The researchers highlighted the importance of real-time social media analysis for improving mental health surveillance and forecasting emotional trends among populations.

3. "Predicting Suicide Risk from Social Media Posts"

Author: John Pestian et al.

Description:

This work investigated the use of natural language processing and sentiment analysis to identify suicide-related expressions in online posts. The proposed approach classified emotional intensity, negative thoughts, and behavioral indicators to estimate suicide risk levels. The study showed that AI-based prediction systems can support early intervention and crisis prevention efforts.

4. "Deep Learning Approaches for Mental Health Prediction"

Author: Yates Andrew et al.

Description:

The authors proposed deep neural network models for detecting mental health conditions using textual and behavioral features from social media platforms. Their system improved prediction accuracy by learning complex emotional patterns and language usage. The research emphasized the effectiveness of deep learning in large-scale mental health forecasting applications.

5. "Twitter-Based Public Health Surveillance System"

Author: Broniatowski David et al.

Description:

This study presented a framework for monitoring public health trends using Twitter data. The system analyzed public discussions, emotional responses, and social interactions to identify emerging health concerns. The findings suggested that social networks can serve as valuable real-time data sources for forecasting national-level health and psychological trends.

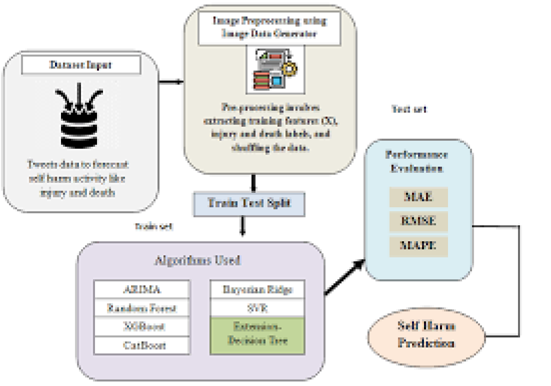
6. "Machine Learning for Early Detection of Self-Harm Behavior"

Author: Sarah E. Valentine et al.

Description:

The research introduced a machine learning-based framework for identifying self-harm behavior from online communication patterns. The model used sentiment analysis, keyword extraction, and behavioral signals to classify high-risk individuals. The study highlighted the importance of predictive analytics in supporting mental health professionals and preventive care systems.

SYSTEM ARCHITECTURE



IMPLEMENTATION

Forecasting National-Level Self-Harm Trends with Social Networks

Now-a-days world is facing a new disease called depression which is the source to cause all diseases even suicides. Peoples may feel depression due to over competition at education and professional levels. Peoples often express depression from their faces or in their writing post skills so author of this paper employing FAST (forecast self-harm patterns) novel technique which analyse users post from online social networking sites and then extract emotions and sentiments from tweets to create training dataset. Author has collected suicides and death data from Thailand health department and then combines both emotions and death values to form a dataset.

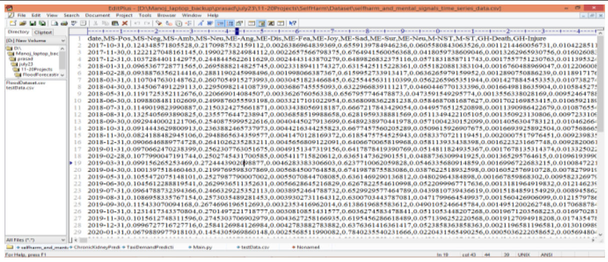
Generated dataset is training with various machine learning algorithms such as ARIMA, SVM, Bayesian Ridge, XGBOOST, Random Forest and CATBOOST to forecast death and injury. In all models ARIMA is showing worst performance and XGBOOST giving best performance.

This trained model can be applied on user's new tweet data to forecast self harm activity like injury and death. This prediction can help government to send all those depressed peoples for counselling to reduce self harm activities.

Author has publish generated dataset and can be download from below link

https://github.com/krittintey/psimilan-fast/blob/main/dataset/selfharm_and_mental_signals_time_series_data.csv

In below screen showing dataset details



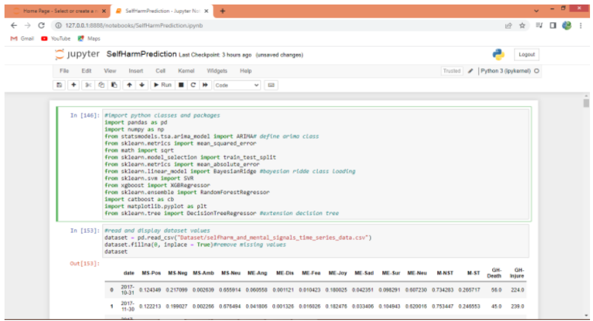
In above dataset screen first row represents dataset column names and remaining rows represents dataset values and each columns contains emotions and sentiments values like positive, negative, neutral etc. So by using above dataset we will train and test all algorithm performance. In propose paper author evaluated each algorithm performance in terms of MAE, RMSE and MAPE metrics. Each metric represents difference between True Test value and predicted value so the lower the difference the better is the algorithm.

Extension Concept

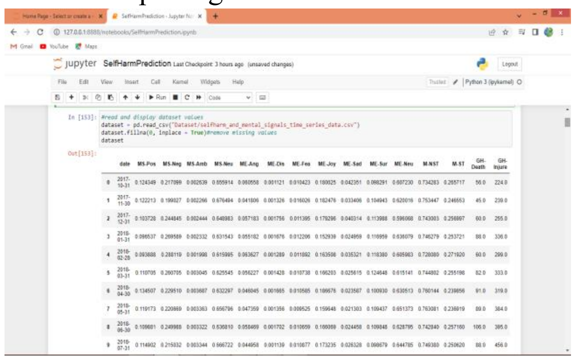
As extension work we have experimented with Decision Tree algorithm which is one of the best algorithm in machine learning and this algorithm is giving lesser MAE error compare to other algorithms.

SCREEN SHOTS

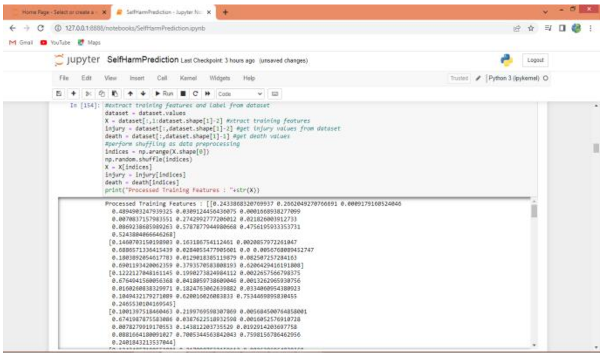
We have coded this project using JUPYTER notebook and below are the code and output screens with blue colour comments



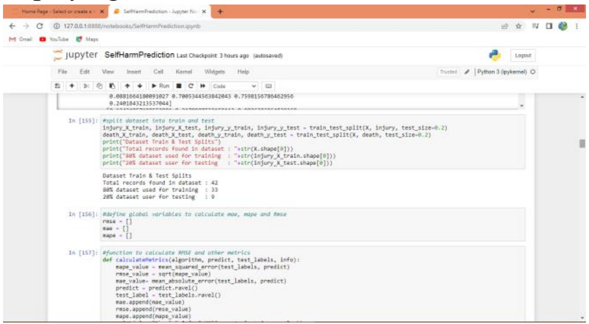
In above screen importing all require python classes and packages



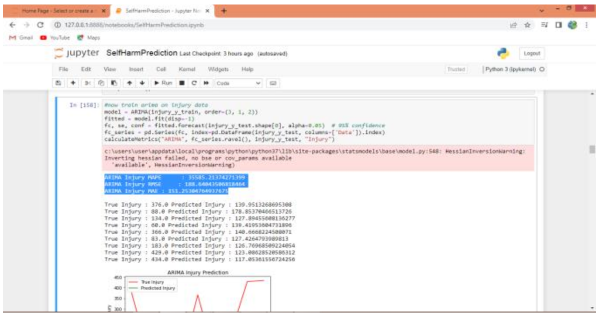
In above screen reading and displaying dataset values



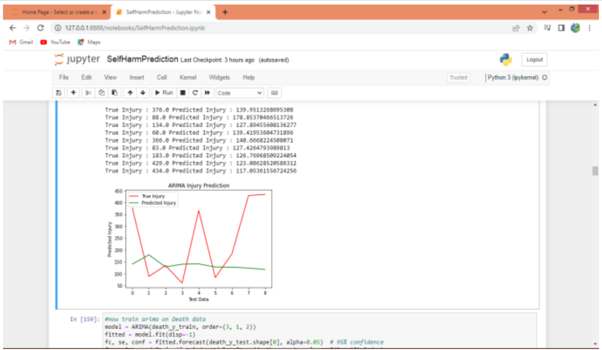
In above screen performing dataset pre-processing like extracting X training features, injury and death labels and then shuffling and displaying all dataset values



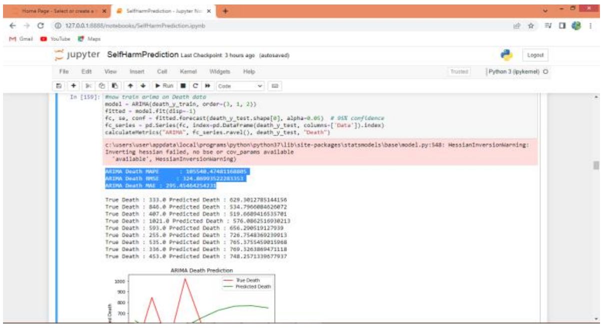
In above screen splitting dataset into train and test where application using 80% dataset for training and 20% for testing and then defining function to calculate MAE, RMSE and MAPE



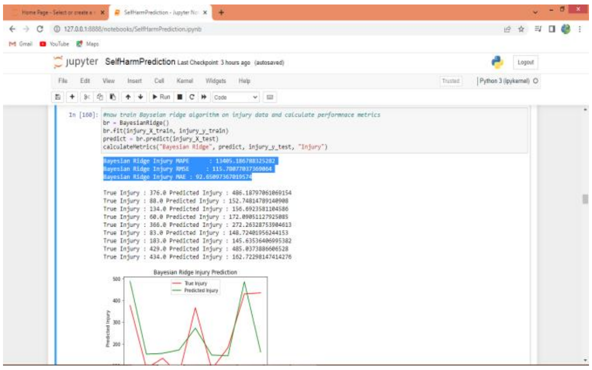
In above screen training ARIMA algorithm on INJURY data and then displaying MAE, RMSE and MAPE metrics in blue colour lines and then displaying True injury and predicted injury values



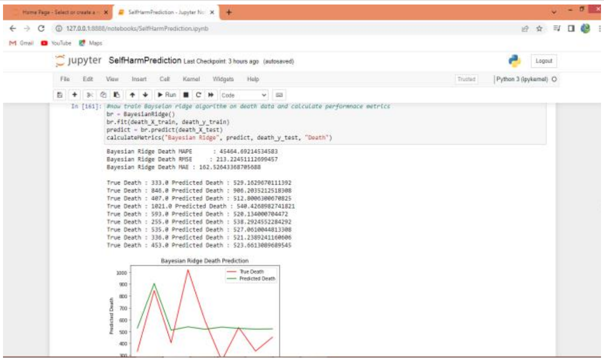
In above graph x-axis represents number of test data and y-axis represents Injury values and red line represents True Injury and green line represents Predicted injury and there is lots of gaps between red and green .line so ARIMA performance is not good. If predicted values are accurate then both lines will overlap.



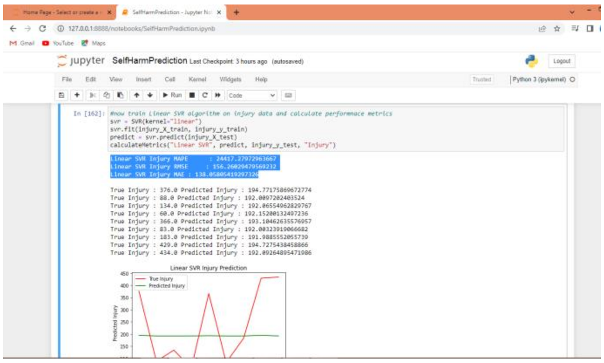
In above screen training ARIMA on DEATH dataset and can see its metrics values in blue colour text



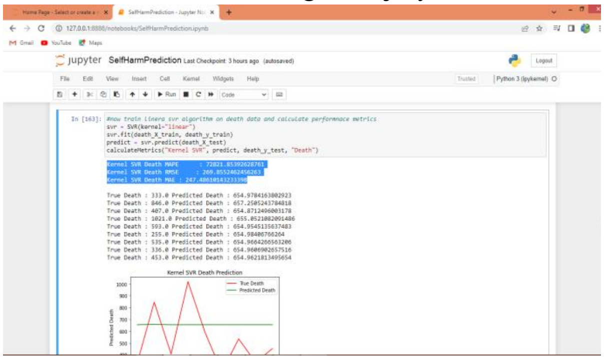
In above screen training Bayesian Ridge on Injury data and then displaying its performance MAE and other values and in graph both lines are overlapping with little gap



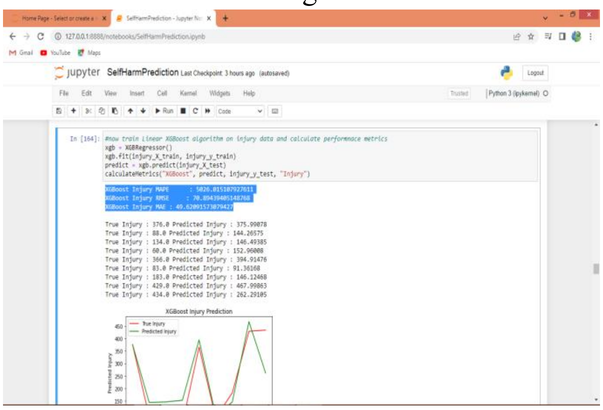
Above is the Bayesian ridge training on death data



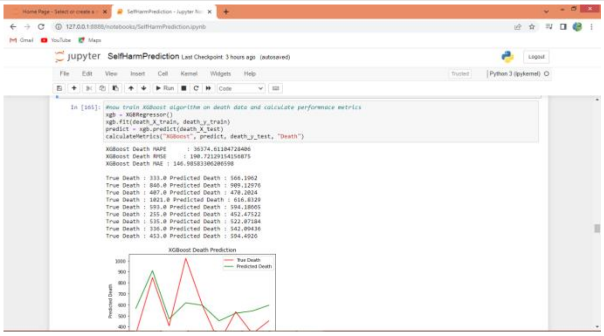
Above is the SVM training on Injury data



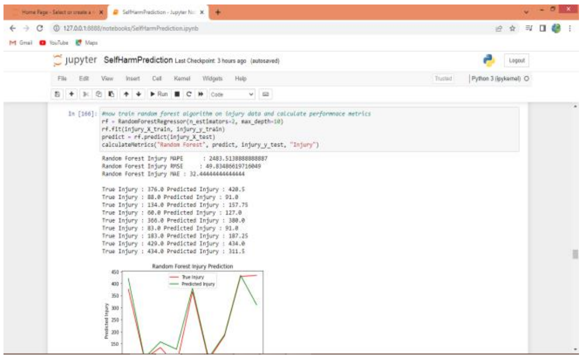
Above is the SVM training on death data



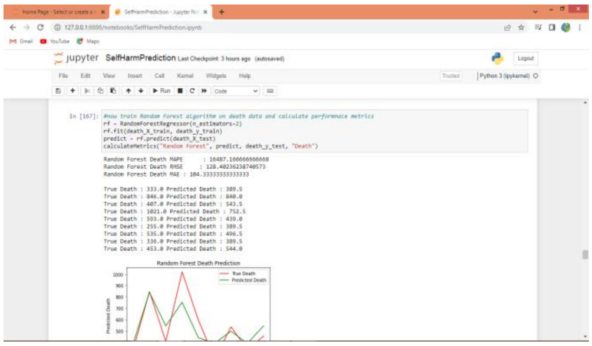
Above is the XGBOOST training on injury data



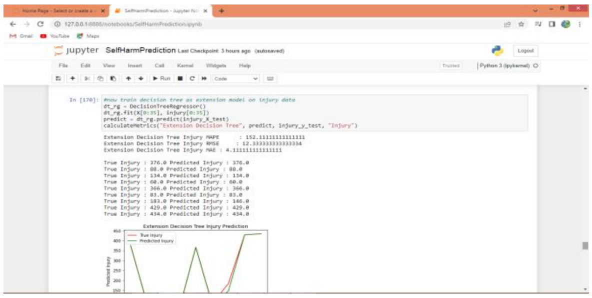
Above is the XGBOOST training on Death data



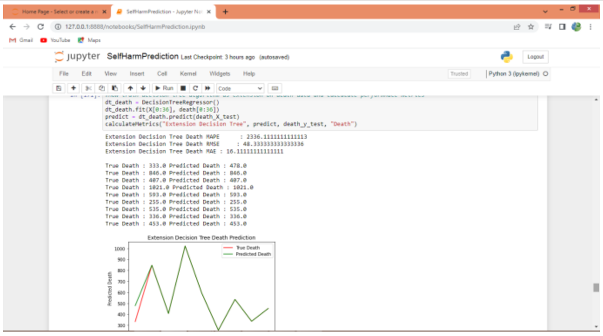
Above is the random forest training on Injury data



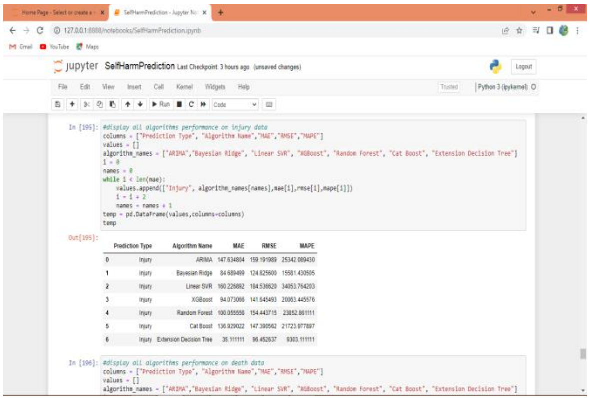
Above is the random forest training on Death data



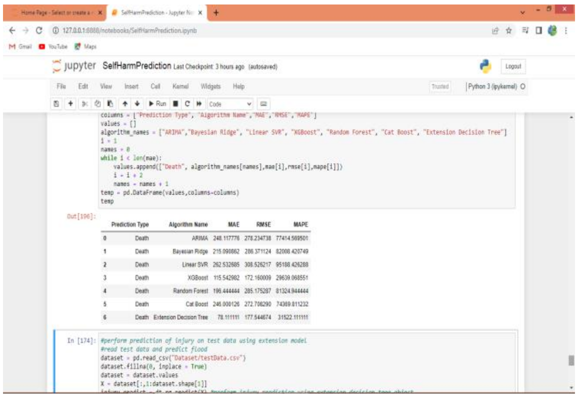
In above screen training extension Decision Tree algorithm on injury data and can see performance metrics



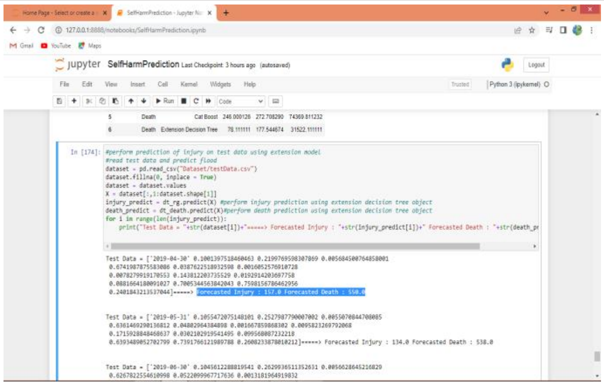
In above screen training extension decision tree on Death data and can see its performance metrics



In above screen can see each algorithm MAE, RMSE and MAPE error values of all algorithms on INJURY data and this metrics are represented as error so if error values are less then algorithm will be consider as best and in all algorithms XGBOOST and extension decision tree got less error rates



In above screen can see each algorithm MAE, RMSE and MAPE error values of all algorithms on DEATH data and in all algorithms XGBOOST and extension decision tree got less error rates



In above screen reading test dataset values and then performing forecasting of INJURY and DEATH and in output before arrow symbol we can see test data values and after arrow symbol can see forecasted values of INJURY and DEATH

CONCLUSION

The proposed system for forecasting national-level self-harm trends using social networks demonstrates the potential of artificial intelligence and social media analytics in improving mental health monitoring and public health decision-making. By analyzing large-scale online interactions, emotional expressions, and behavioral patterns, the framework can identify early warning signals related to self-harm and psychological distress. The integration of machine learning, natural language processing, and predictive analytics enhances the accuracy and efficiency of forecasting compared to traditional monitoring approaches.

The study highlights how real-time social network data can support healthcare organizations, policymakers, and mental health professionals in planning timely interventions, awareness campaigns, and resource allocation strategies. The proposed approach also emphasizes ethical data handling, privacy protection, and responsible use of publicly available information. Overall, the system contributes to the development of intelligent and proactive mental health support systems capable of addressing growing psychological health challenges at a national scale.

FUTURE WORK

Future enhancements of the proposed system can focus on improving prediction accuracy, scalability, and real-time monitoring capabilities. Advanced deep learning models and transformer-based natural language processing techniques can be integrated to better understand complex emotional expressions, multilingual content, and contextual meanings in social media posts. The system can also be expanded to analyze data from multiple social networking platforms, online forums, and digital communities to provide more comprehensive mental health insights.

Further research may include the incorporation of multimedia analysis such as images, videos, and voice data to detect emotional and behavioral indicators more effectively. Real-time alert mechanisms and automated intervention support systems can also be developed to assist healthcare professionals and mental health organizations in responding quickly to high-risk situations. Additionally, integrating geographical and demographic analysis may help identify region-specific mental health trends and improve public health planning.

Future work should also focus on strengthening ethical standards, privacy preservation, and bias reduction in AI models to ensure responsible use of sensitive social data. Collaboration with healthcare institutions and government agencies can further improve the reliability and practical implementation of national-level self-harm forecasting systems.

REFERENCE

1. Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz, “Predicting Depression via Social Media,” *Proceedings of the International AAAI Conference on Web and Social Media*, 2013.
2. Glen Coppersmith, Mark Dredze, and Craig Harman, “Quantifying Mental Health Signals in Twitter,” *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology*, 2014.
3. John Pestian, Pawel Matykiewicz, and Amy Linn-Gust, “Suicide Note Classification Using Natural Language Processing,” *Biomedical Informatics Insights*, 2012.
4. Andrew Yates, Arman Cohan, and Nazli Goharian, “Depression and Self-Harm Risk Assessment in Online Forums,” *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017.
5. Kavuri, S. (Ed.). (2024). Shift-left and shift-right testing approaches: A practical roadmap for continuous quality in agile and DevOps. Journal of Information Systems Engineering and Management, 9(4). https://doi.org/10.52783/jisem.v9i4.127
6. Gajula, S. (2026). Two Pillars of Banking Intelligence: A Comparative Analysis of AI Techniques for Fraud Prevention and Churn Mitigation. 2026 14th International Symposium on Digital Forensics and Security (ISDFS), 1–6. https://doi.org/10.1109/isdfs69419.2026.11458995
7. Gummadi, V. P. K., Chilamkurthi, L. S., & Kavuri, S. (2026). Securing APIs in Government Clouds and Runtime Fabric Using FIPS-Enabled MuleSoft. 2026 International Conference on Artificial Intelligence, Systems, and Emerging Technologies (ICAISET), 1–6. https://doi.org/10.1109/icaiset66439.2026.11542099
8. Mudusu, S. K. (2022). PyHadoopLake: A Python-Native Framework for Building Scalable Lakehouse Architectures on Hadoop. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 5(5), 7449-7452.
9. Pavan Kumar Adabala. (2026). Smart Retail Fuel Systems: IoT-Enabled Solutions for Loss Prevention and

Environmental Safety. Computer Fraud and Security, 868–875. https://doi.org/10.52710/cfs.995

10. Maturi, S. Y. Cryptographic Privacy Engines: Practical Multi-Party Protocols For Confidential Database Queries.
11. David A. Broniatowski, Michael J. Paul, and Mark Dredze, "National and Local Influenza Surveillance through Twitter," *PLOS ONE*, 2013.
12. Sarah E. Valentine, Kelly M. Smith, and colleagues, "Machine Learning Approaches for Detecting Self-Harm Behaviors in Social Media," *Journal of Medical Internet Research*, 2019.
13. Manoharan, D. (2026). Advancing Healthcare EDI Interoperability Through Informatica Cloud B2B Gateway Quality Engineering. Available at SSRN 6385719.
14. Kumar Adabala, P. (2021). Optimizing ERP Modernization: A Smart Data Migration Framework Approach. International Journal of Enhanced Research in Science, Technology &Amp; Engineering, 10(07), 61–72. https://doi.org/10.55948/ijerste.2021.0708
15. Michael J. Paul and Mark Dredze, "You Are What You Tweet: Analyzing Twitter for Public Health," *Proceedings of the International AAAI Conference on Web and Social Media*, 2011.
16. H. A. Schwartz, Johannes Eichstaedt, and Margaret L. Kern, "Personality, Gender, and Age in the Language of Social Media," *PLOS ONE*, 2013.