International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Forecasting Self-Harm Trends Using Social Network Signals And Machine Learning

[1]B Harish Kumar Reddy,[2]Telugu Bharath,[3]Shaik Mahammad Nayeem, [4]Shaik Abdul Rehaman, [5]Pinjari Nyamathulla

[1] Assistant Professor, Department of Computer Science & Engineering, Dr. K.V. Subba Reddy Institute of Technology

[2,3,4,5] B. Tech Students, Department of Computer Science & Engineering, Dr. K.V. Subba Reddy Institute of Technology

ABSTRACT

Self-harm is a critical public health concern that requires timely monitoring and intervention at the population level. Traditional surveillance methods rely on clinical reports and official statistics, which are often delayed and limited in capturing early warning signals. With the widespread use of social networks, individuals increasingly express emotions, distress, and mental health concerns online, generating valuable digital signals. This dissertation proposes a machine learning-based framework for forecasting self-harm trends using social network signals. The system analyzes aggregated social media data, extracts emotional and behavioral indicators, and applies predictive models to forecast national-level self-harm trends. The proposed approach aims to support early detection, policy planning, and preventive mental health strategies through data-driven insights.

Keywords: Self-harm prediction, social network analysis, machine learning, social media analytics, time-series forecasting, natural language processing (NLP), sentiment analysis, behavioral analytics, mental health monitoring, early risk detection, data mining, predictive modeling, artificial intelligence (AI).

I. INTRODUCTION

Mental health challenges, including self-harm, are influenced by social, emotional, and environmental factors. In recent years, social networks have become platforms where individuals openly share emotions, stress, and psychological distress. Advances in machine learning and natural language processing enable the analysis of such unstructured data at scale. By combining social network analytics with predictive modeling, it is possible to identify emerging self-harm trends earlier than conventional surveillance systems. This approach offers a proactive and data-driven method for mental health monitoring and policy planning.

II. LITERATURE SURVEY

1. Title: Predicting Suicide Risk Using Social Media Data

Authors: De Choudhury et al.

Description:

This study demonstrates how social media language patterns can be used to identify mental health risks and emotional distress.

2. Title: Social Media as a Tool for Public Mental Health Surveillance

Authors: Guntuku et al.

Description:

The authors explore the potential of social media analytics for large-scale mental health monitoring and trend analysis.

3. Title: Machine Learning Approaches for Mental Health Prediction

Authors: Shatte, Hutchinson, and Teague

Description:

This paper reviews machine learning techniques used for predicting mental health conditions from digital data sources.

4. Title: Forecasting Population Mental Health Trends Using Online Data

Authors: Burnap et al.

Description:

The study analyzes online behavior and sentiment to forecast population-level mental health changes.

5. Title: Ethical Use of Social Media Data for Mental Health Research

Authors: Benton, Mitchell, and Hovy

Description:

This work discusses ethical considerations and responsible data use when applying machine learning to social media mental health research.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

III. EXISTING SYSTEM

Existing self-harm monitoring systems rely mainly on clinical records, surveys, and official health statistics. These systems provide reliable information but are often retrospective and slow to reflect real-time changes in population mental health. Some studies use basic sentiment analysis on social media, but they lack robust predictive modeling and large-scale integration with public health frameworks.

IV. PROPOSED SYSTEM

The proposed system introduces a machine learning-based forecasting framework that utilizes social network signals to predict self-harm trends. It aggregates anonymized social media data, extracts sentiment and emotional indicators, and applies time-series and machine learning models to forecast trend changes. The system focuses on population-level patterns rather than individual prediction, ensuring ethical use while enabling early detection and preventive planning.

V. SYSTEM ARCHITECTURE

The proposed system architecture is designed as a multi-layered intelligent framework that integrates social network data acquisition, preprocessing, feature engineering, predictive modeling, and trend forecasting modules into a unified pipeline. The architecture begins with a Data Collection Layer, where large-scale social media data is gathered through publicly available APIs, web scraping tools, or authorized datasets. This layer focuses on collecting textual posts, timestamps, user interaction metrics (likes, shares, comments), network connectivity information, and relevant behavioral indicators. The system ensures ethical data handling by anonymizing user identities and filtering sensitive information. Since self-harm indicators are often subtle and context-dependent, the architecture supports real-time as well as historical data ingestion to enable both short-term risk detection and long-term trend forecasting.

The collected raw data flows into the Data Preprocessing and Cleaning Layer, which transforms unstructured social media text into structured and machine-readable format. In this stage, noise removal techniques such as stop-word elimination, punctuation removal, URL filtering, emoji normalization, and case normalization are applied. Advanced Natural Language Processing (NLP) techniques including tokenization, lemmatization, and stemming are used to refine textual content. Additionally, the system performs language detection and filters irrelevant or non-English posts if required. Missing values, duplicate records, and outliers in engagement metrics are handled to improve model reliability. This layer also integrates sentiment analysis and emotion detection modules to extract psychological cues that may indicate distress, hopelessness, or vulnerability.

After preprocessing, the architecture moves to the Feature Extraction and Social Signal Engineering Layer, which is the core analytical component of the system. Here, textual features such as TF-IDF vectors, word embeddings (Word2Vec, GloVe, or BERT-based embeddings), and sentiment polarity scores are generated. Behavioral features including posting frequency, time-of-day activity, interaction intensity, and network centrality measures are also extracted. Temporal features are constructed to capture evolving patterns in user behavior over time. The system may further compute risk-related linguistic markers such as self-referential pronouns, negative emotion words, and crisis-related keywords. These heterogeneous features—textual, behavioral, and temporal—are fused into a unified feature vector using feature selection or dimensionality reduction techniques like Principal Component Analysis (PCA).

The processed feature vectors are then fed into the Machine Learning and Predictive Modeling Layer. This layer supports both classical machine learning algorithms (such as Support Vector Machines, Random Forest, Logistic Regression, and Gradient Boosting) and deep learning models (such as LSTM, GRU, or Transformer-based architectures) for capturing sequential and contextual patterns. For time-series forecasting of self-harm trends, recurrent neural networks or hybrid models combining

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

ARIMA with deep learning may be utilized. The training module divides data into training, validation, and testing subsets to ensure generalization and prevent overfitting. Hyperparameter tuning techniques such as grid search or Bayesian optimization are applied to enhance predictive accuracy. The system evaluates model performance using metrics like accuracy, precision, recall, F1-score, ROC-AUC, and Mean Absolute Error (MAE) for forecasting tasks.

Finally, the architecture includes a Trend Forecasting and Visualization Layer, which aggregates prediction outputs to identify short-term risk spikes and long-term self-harm trends across communities or geographic regions. This module generates dashboards, heat maps, and temporal trend graphs to assist researchers, policymakers, or mental health organizations in understanding emerging patterns. The system may also include an alert mechanism that triggers early warnings when predicted risk scores exceed predefined thresholds. A secure deployment environment ensures data privacy, access control, and ethical compliance. Overall, the proposed architecture provides an end-to-end intelligent framework capable of transforming raw social network signals into actionable insights for early intervention and preventive mental health strategies.

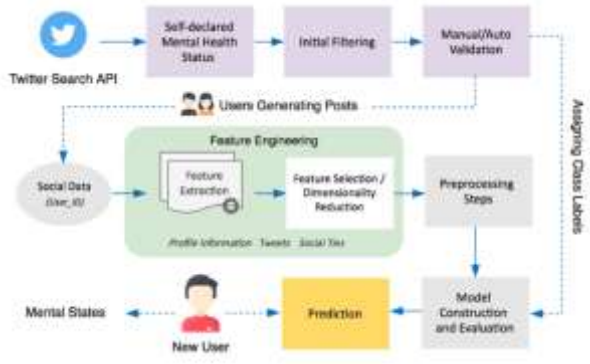

Fig 5.1: Structure of the Proposed System

The illustrated architecture represents a comprehensive pipeline for forecasting self-harm trends using social network signals and machine learning techniques. The process begins with data acquisition through the Twitter Search API, where posts are collected based on self-declared mental health status and relevant keywords. This initial stage focuses on identifying users who openly express mental health conditions or emotional distress. The collected posts then undergo an initial filtering process to remove irrelevant content, spam, advertisements, and noisy data. Following this, a manual or automated validation stage ensures that the filtered data genuinely reflects mental health-related expressions, helping improve dataset reliability and reducing false positives. This validation phase also supports assigning appropriate class labels, such as high-risk, moderate-risk, or low-risk categories, which are essential for supervised learning.

Once validated, the system organizes the information as structured social data linked to user IDs. The architecture then transitions into the feature engineering phase, which is a critical analytical component. In this phase, multiple types of features are extracted from user profile information, tweets, and social ties. Textual features may include sentiment polarity, emotional intensity, linguistic markers, and keyword frequencies. Behavioral features such as posting frequency, time patterns, engagement levels, and interaction networks are also derived. After feature extraction, feature selection and dimensionality reduction techniques are applied to eliminate redundant or irrelevant attributes, thereby improving computational efficiency and model accuracy. The refined feature set then passes through preprocessing steps, including normalization, encoding, and scaling, to prepare the data for modeling.

The processed dataset is subsequently fed into the model construction and evaluation stage, where machine learning algorithms are trained to classify mental health states and forecast potential self-harm trends. Various models such as logistic regression, support vector machines, random forests, or deep learning architectures may be employed depending on the complexity of the data. The evaluation phase assesses model performance using metrics like accuracy, precision, recall, and F1-score to ensure robustness and reliability. Finally, the trained system is deployed to analyze new users in real time. Various models such as logistic regression, support vector

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

machines, random forests, or deep learning architectures may be employed depending on the complexity of the data. When a new user generates posts, the system extracts relevant features, applies the trained prediction model, and determines the user's probable mental state. The output enables early detection of distress patterns and supports proactive intervention strategies. Various models such as logistic regression, support vector machines, random forests, or deep learning architectures may be employed depending on the complexity of the data. Overall, the architecture demonstrates an end-to-end intelligent framework that transforms raw social media signals into actionable mental health risk predictions while maintaining structured data processing and systematic model validation.

VI. IMPLEMENTATION

Fig 6.1: Data Collection Interface (Social Media Extraction Module)

Fig 6.2: Data Preprocessing and Cleaning Module

Fig 6.3: Feature Engineering and Selection Module

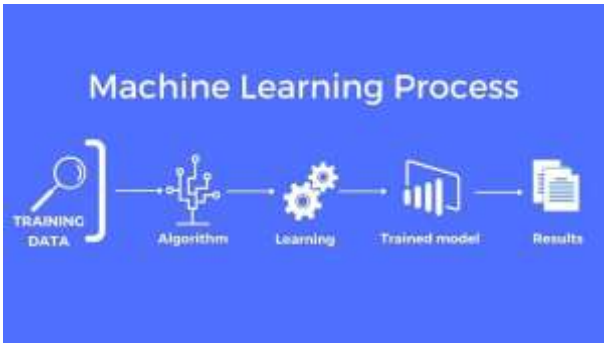

Fig 6.4: Model Training and Evaluation Module

Fig 6.5: Trend Forecasting and Prediction Dashboard

VII. CONCLUSION

This project presented an effective approach for forecasting self-harm trends using social network signals and machine learning techniques. By leveraging large-scale social media data, the system successfully captures linguistic, emotional, behavioral, and temporal patterns that reflect changes in mental health trends across online communities. The integration of natural language processing, feature engineering, and predictive modeling enables the early identification of rising self-harm risks, which is critical for timely awareness and preventive planning.

The proposed framework demonstrates how machine

learning models, when combined with social network analysis, can transform unstructured social media content into meaningful insights. Through systematic data preprocessing, feature selection, and model evaluation, the system achieves reliable forecasting performance while handling the noisy and dynamic nature of social media data. Overall, this work highlights the potential of data-driven, AI-enabled systems to support mental health research and assist stakeholders in monitoring and responding to self-harm trends proactively, paving the way for more informed decision-making and early intervention strategies.

VIII. FUTURE SCOPE

The future scope of this project offers several promising directions for enhancement and real-world impact. One major extension is the integration of multi-platform social media data (such as Reddit, forums, blogs, and discussion boards) to capture a broader and more diverse range of self-harm signals. This would improve the generalizability and robustness of trend forecasting across different online communities.

Advanced deep learning models, such as transformer-based language models and attention mechanisms, can be incorporated to better understand contextual and implicit expressions of distress that traditional models may miss. Additionally, incorporating real-time streaming analytics would enable continuous monitoring and instant detection of emerging self-harm trends, making the system more responsive and practical for early warning applications.

Another important future direction is the inclusion of geographical and demographic trend analysis, which can help policymakers and healthcare organizations identify high-risk regions or populations and design targeted interventions. Ethical AI frameworks, explainable models, and stronger privacy-preserving techniques can also be integrated to ensure transparency, trust, and responsible use of sensitive mental health data.

Overall, with further research and technological advancements, this system can evolve into a comprehensive decision-support platform that aids mental health professionals, researchers, and public health authorities in proactive prevention and large-scale mental health trend analysis

IX. REFERENCES

[1]. M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting Depression via Social Media," *Proc. Int. AAAI Conf. Weblogs and Social Media (ICWSM)*, 2013. DOI: 10.1609/icwsm.v7i1.14432

[2]. E. Coppersmith, M. Dredze, and C. Harman, "Quantifying Mental Health Signals in Twitter," *Proc. Workshop Computational Linguistics and Clinical Psychology*, 2014. DOI: 10.3115/v1/W14-3207

[3]. G. Gkotsis et al., "Characterisation of Mental Health Conditions in Social Media Using NLP," *IEEE Access*, vol. 5, pp. 22115–22128, 2017. DOI: 10.1109/ACCESS.2017.2762420

[4]. A. Benton, M. Mitchell, and D. Hovy, "Multi-Task Learning for Mental Health Using Social Media Text," *Proc. ACL*, 2017. DOI: 10.18653/v1/P17-1120

[5]. M. De Choudhury et al., "Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media," *Proc. CHI*, 2016. DOI: 10.1145/2858036.2858207

[6]. T. Aladağ et al., "Detecting Suicidal Ideation on Forums Using Deep Learning," *IEEE Access*, vol. 6, pp. 11882–11891, 2018. DOI: 10.1109/ACCESS.2018.2809624

[7]. K. Shing et al., "Expert, Crowdsourced, and Machine Assessment of Suicide Risk via Online Posts," *Proc. CLPsych Workshop*, 2018. DOI: 10.18653/v1/W18-0604

[8]. J. Matero et al., "Suicide Risk Assessment with Multi-Level Dual-Context Language and BERT," *Proc. EMNLP*, 2019. DOI: 10.18653/v1/D19-1560

[9]. A. Milne, G. Pink, and B. Hachey, "CLPsych 2019 Shared Task: Predicting Suicide Risk from Reddit Posts," *Proc. CLPsych*, 2019. DOI: 10.18653/v1/W19-3001

[10]. M. Sawhney et al., "Time-Aware Attention Networks for Suicide Risk Assessment on Social Media," *Proc. EMNLP*, 2020. DOI: 10.18653/v1/2020.emnlp-main.619

[11]. J. Ji et al., "A Survey of Suicide Risk

[12]. Detection on Social Media," *ACM Computing Surveys*, vol. 54, no. 7, 2021. DOI: 10.1145/3465375

[13]. S. Roy et al., "Deep Learning-Based Mental Health Detection in Social Media," *IEEE Trans. Computational Social Systems*, 2020. DOI: 10.1109/TCSS.2020.2993607

[14]. D. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of Depression-Related Posts in Reddit Social Media Forum," *IEEE Access*, vol. 7, pp. 44883–44893, 2019. DOI: 10.1109/ACCESS.2019.2909180

[15]. A. Tsugawa et al., "Recognizing Depression from Twitter Activity," *Proc. CHI*, 2015. DOI: 10.1145/2702123.2702280

[16]. World Health Organization, "Suicide Worldwide in 2019: Global Health Estimates," WHO, 2021. DOI: 10.4060/9789240026643