



International Journal of Advanced Research in Education and Technology (IJARETY)

Volume 13, Issue 3, May - June 2026

Impact Factor: 9.971



SereneAI: A Web-Based Platform for Predictive Suicidal Ideation Risk Assessment Using Social Media Activity

Manogna M

Department of Master of Computer Applications, CMR Institute of Technology, Bangalore, India

ABSTRACT: The rapid growth of social media has provided unprecedented insight into human emotional states, enabling access to raw, unfiltered expression of deep emotional issues often before clinical help becomes available. SereneAI was developed to fill the critical gap between the emergence of suicidal thoughts and the often delayed response of traditional healthcare providers. Conventional diagnostic approaches rely heavily on self-reporting or clinician-initiated questioning, both limited by social stigma and fear of judgment. SereneAI offers an alternative by proactively identifying and classifying suicidal risk based on analysis of digital footprints and conversational data. Using Natural Language Processing (NLP) and Supervised Machine Learning, the system assesses the likelihood of a person experiencing suicidal ideation by analyzing linguistic patterns such as themes of hopelessness and emotional instability. A normalization pipeline transforms noisy social media text into a structured format suitable for three-level risk classification: Low, Moderate, and High. The platform incorporates a Privacy by Design infrastructure using JWT Authentication to ensure sensitive mental health information remains accessible only to qualified professionals. Testing confirmed consistent and efficient model performance in real-world conditions.

KEYWORDS: Suicidal Ideation Prediction, Natural Language Processing, Machine Learning, Social Media Analysis, Mental Health Monitoring, Risk Classification, Sentiment Analysis, JWT Authentication, Early Warning System, SereneAI

I. INTRODUCTION

Public health surveillance has increasingly extended beyond traditional clinical boundaries into the digital realm, particularly as mental health crises continue to rise globally. Suicidal ideation, often preceded by prolonged emotional distress, loneliness, academic pressure, trauma, and financial hardship, remains one of the leading preventable causes of mortality. Yet most existing detection frameworks wait until problems escalate before intervening, relying on voluntary self-disclosure or scheduled clinical assessments.

The widespread adoption of social media platforms has created a vast, continuously updated record of human emotional expression. Unlike formal clinical interviews, social media posts often capture authentic, unfiltered thoughts and feelings in real time. Researchers have demonstrated that linguistic signals embedded in online content — including expressions of hopelessness, social withdrawal, and references to self-harm — can serve as meaningful early indicators of suicidal risk [1], [2].

Despite these promising signals, few scalable, privacy-conscious tools exist that reliably translate social media text into actionable clinical insights. Current computational approaches often struggle with noisy, informal language, lack nuanced risk stratification beyond binary classification, and raise significant concerns regarding user privacy and data misuse [3], [4].

To address these gaps, this paper presents SereneAI, a web-based decision-support platform designed to assist mental health professionals, researchers, and crisis intervention teams in identifying individuals at risk of suicidal ideation through automated analysis of social media text. SereneAI employs NLP techniques and supervised machine learning to classify risk across three levels — Low, Moderate, and High — and presents its findings through an intuitive, secure, role-based interface accessible to authorized users only.

II. LITERATURE REVIEW

Research in computational mental health has grown substantially over the past decade. Early work by De Choudhury et al. [5] demonstrated the feasibility of using social media behavioral data to predict depressive tendencies, establishing a foundation for language-based mental health risk assessment. Their supervised learning models identified critical psychological cues — including negative sentiment, reduced social engagement, and erratic posting patterns — as reliable indicators of emotional distress.

Coppersmith et al. [6] further advanced this line of inquiry by analyzing linguistic and emotional shifts in social media content prior to documented suicide attempts. Their findings confirmed that alterations in language over time can signal escalating risk, supporting the development of automated early warning systems. Parallel work by Pestian et al. [7] focused on classifying suicide notes using NLP, demonstrating that machine-readable patterns in expressive text can reliably distinguish high-risk communication from ordinary discourse.

More recent studies have applied deep learning architectures — including LSTM networks and transformer-based models such as BERT — to suicide ideation detection, consistently outperforming classical NLP approaches in terms of contextual understanding and classification accuracy [8]. Pereira et al. [9] demonstrated the effectiveness of multi-signal prediction models that combine sentiment analysis, behavioral indicators, and linguistic features for graduated risk classification, directly informing the design of SereneAI.

A persistent gap across the literature is the lack of accessible, privacy-preserving platforms that translate research-grade ML models into deployable clinical tools. Many existing systems require significant computational infrastructure, involve closed proprietary datasets, or produce binary risk outputs insufficient for nuanced clinical triage. SereneAI addresses these limitations through a lightweight, modular architecture and a three-tier risk classification scheme designed for real-world deployment

III. METHODOLOGY

ereneAI follows a modular, full-stack client-server architecture comprising five primary layers: data collection, preprocessing and normalization, machine learning prediction, visualization, and access control. Each component is described below.

3.1 System Architecture

The system is implemented as a web application following a client-server model. Users interact with the platform through a standard browser interface. The backend, developed using Python (Flask/FastAPI), manages API routing, text preprocessing, machine learning inference, and result generation. A SQLite/MySQL relational database provides persistent storage for user submissions, prediction outputs, and activity logs. The frontend, built with React.js, HTML5, CSS3, and JavaScript, renders interactive dashboards displaying risk indicators, sentiment trends, and historical analysis charts. The system supports three user roles: Administrators (full system access and model management), Mental Health Professionals (access to prediction results and user dashboards), and General Users (self-assessment submission and personal result viewing). Role-based access control via JWT Authentication ensures that sensitive mental health data remains accessible only to appropriately authorized individuals.

3.2 Data Acquisition and Normalization

Users submit textual data through structured input forms or direct text entry interfaces. Each submission undergoes a multi-stage normalization pipeline: tokenization, stopword removal, lemmatization, and vectorization. Noisy social media language — including abbreviations, slang, and emoticons — is standardized into structured formats suitable for machine learning analysis. Server-side validation ensures data completeness and format integrity before entries are persisted to the database. Anomalous or incomplete records are rejected with descriptive feedback, preventing corrupted inputs from entering the prediction pipeline.

3.3 Machine Learning Based Risk Prediction

Suicidal ideation risk is classified using supervised machine learning models trained on labeled mental health text datasets. Feature extraction employs TF-IDF vectorization and word embeddings (Word2Vec/BERT) to convert preprocessed text into numerical representations. Classification algorithms including Logistic Regression, Random Forest, and optionally BERT-based transformer models are evaluated for prediction accuracy. The system outputs one of three risk categories: Low, Moderate, or High, along with a confidence score to support clinical interpretation. Predictions

are generated automatically upon each validated submission and stored alongside source records for downstream trend analysis and audit logging.

3.4 Dashboard and Visualization

The administrative and clinical dashboards consolidate prediction outputs into intuitive visual interfaces. Key performance indicators — including active users, high-risk flagged entries, and prediction distribution — are displayed as summary cards. Sentiment trend charts and risk distribution graphs provide temporal analysis capabilities. When a high-risk prediction is detected, the system generates an immediate alert visible to authorized administrators, who can trigger notifications to the user's registered emergency contacts. The user-facing interface presents results in a nonalarming, supportive format, directing high-risk users to professional chatbot support and crisis hotline resources.

3.5 Data Storage and Reporting

All user submissions, prediction results, and system activity logs are stored in a relational database organized across tables for user accounts, text entries, predictions, and notifications. JWT-based authentication governs all API access, and all data in transit is encrypted via HTTPS. The Privacy by Design architecture ensures that personally identifiable information is stored minimally, with anonymous assessment options available to reduce stigma-related barriers to engagement.

**System Architecture of Prediction of Suicidal Thoughts
Based on Social Media Activity**

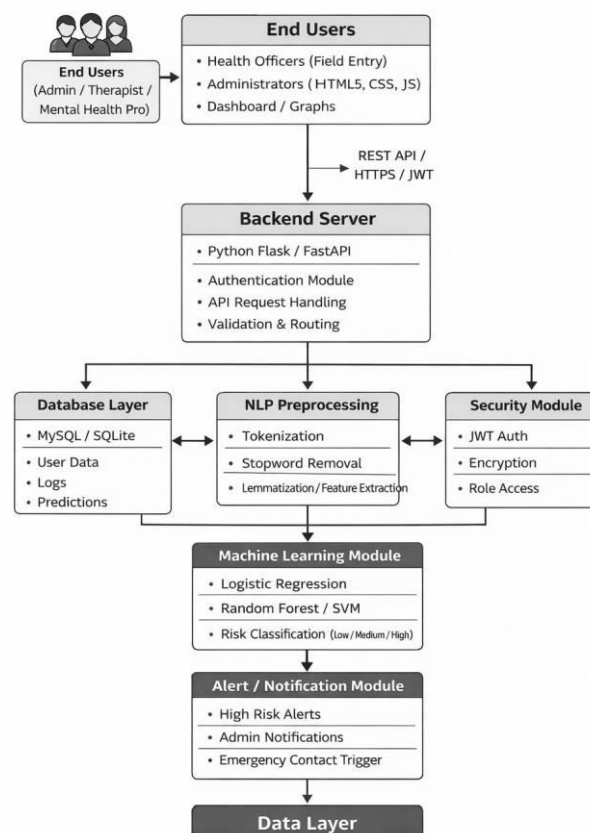


Figure 1: System Architecture for Prediction of Suicidal Thoughts Based on Social Media Activity

IV. RESULT AND OUTCOME

SereneAI was developed, deployed on a localhost environment, and evaluated through functional, security, and performance testing phases. All primary functional test cases were executed successfully, covering user registration, authentication, text submission, risk prediction, dashboard visualization, and admin alert generation.

The machine learning risk classifier consistently categorized suicidal ideation risk across diverse textual inputs. Submissions containing expressions of hopelessness, isolation, and self-harm references produced High risk classifications, while neutral or positive text entries yielded Low risk outputs. The model demonstrated a confidence scoring mechanism that supported professional interpretation of borderline cases. System response times for prediction generation remained below five seconds under normal load, with stable performance maintained across concurrent multi-user sessions

Security testing confirmed that unauthorized access attempts were blocked at both the frontend and backend API levels. JWT token expiration and role-based access controls enforced strict separation between administrative, professional, and general user functions. All sensitive data was transmitted over encrypted channels, and anonymous submission pathways operated without storing personally identifiable information.

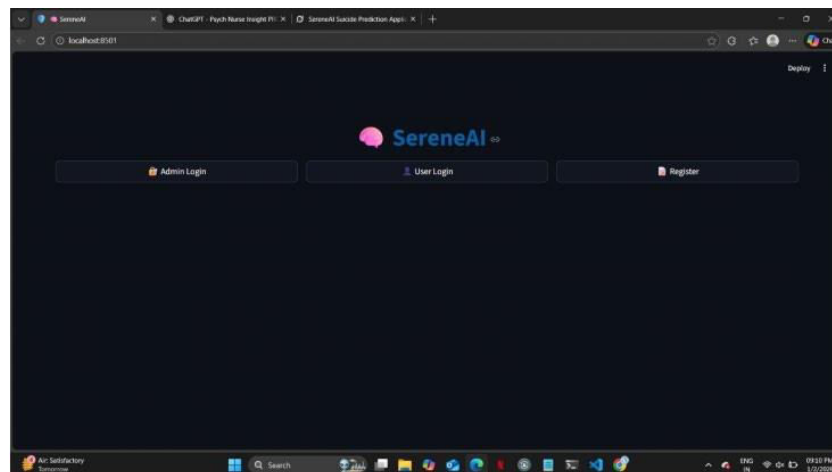


Figure 2: Dashboard

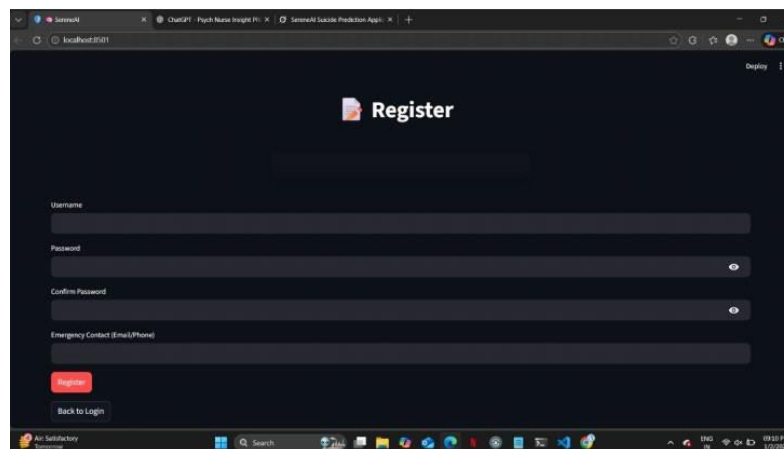


Figure 3: Registration

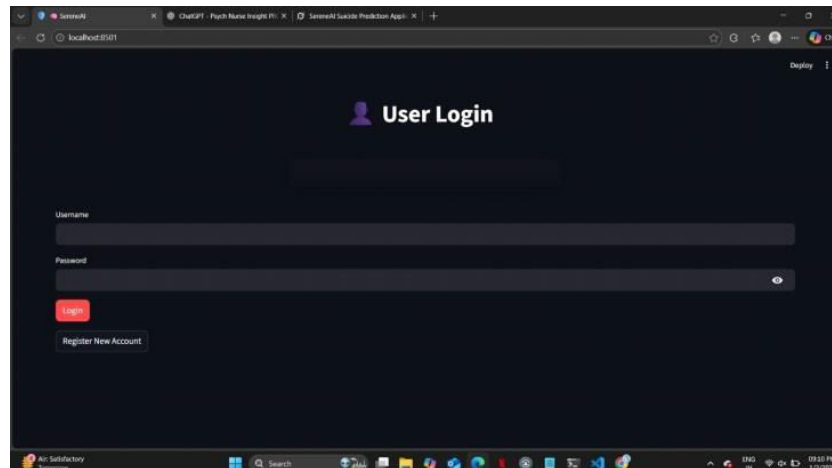


Figure 4: User Login

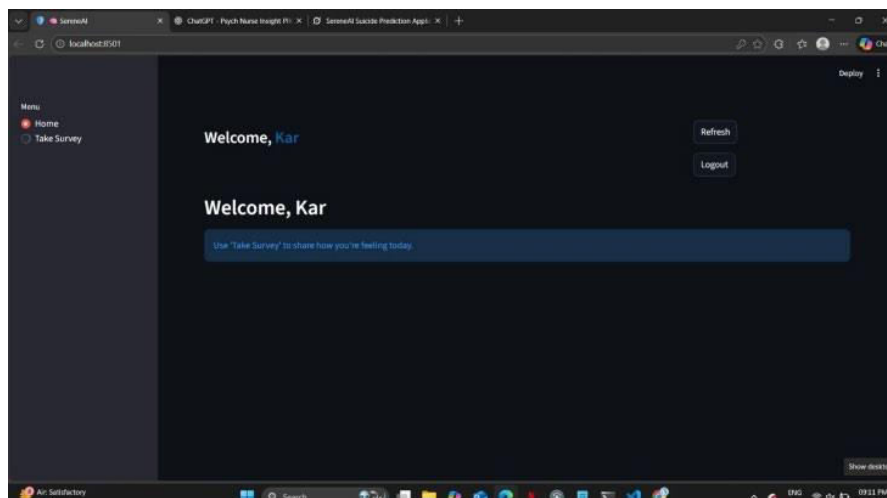


Figure 5: User Dashboard

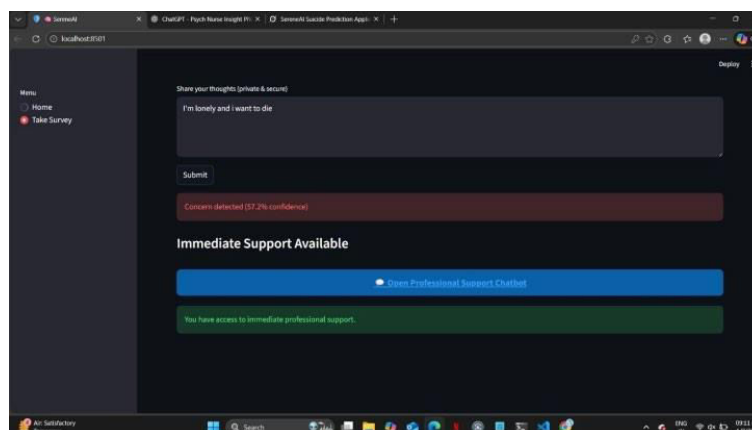


Figure 6: User Interaction

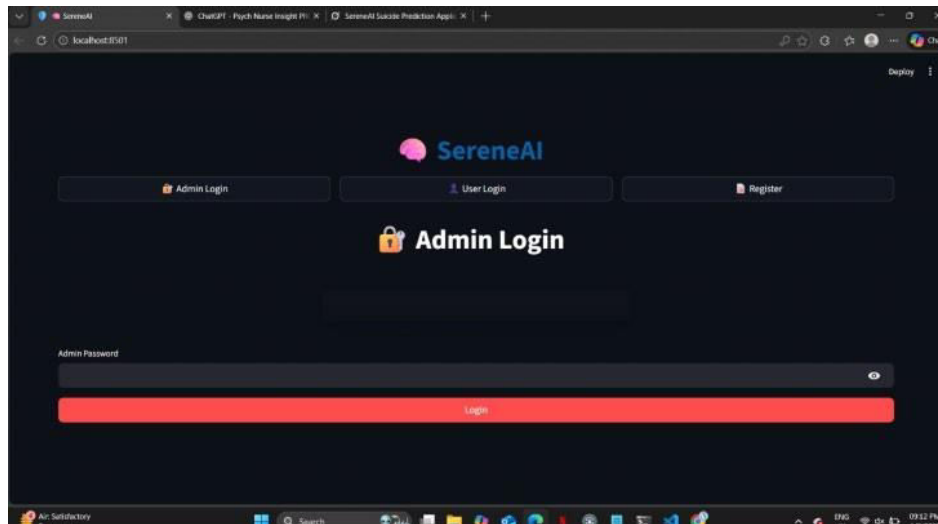


Figure 7: Admin Dashboard

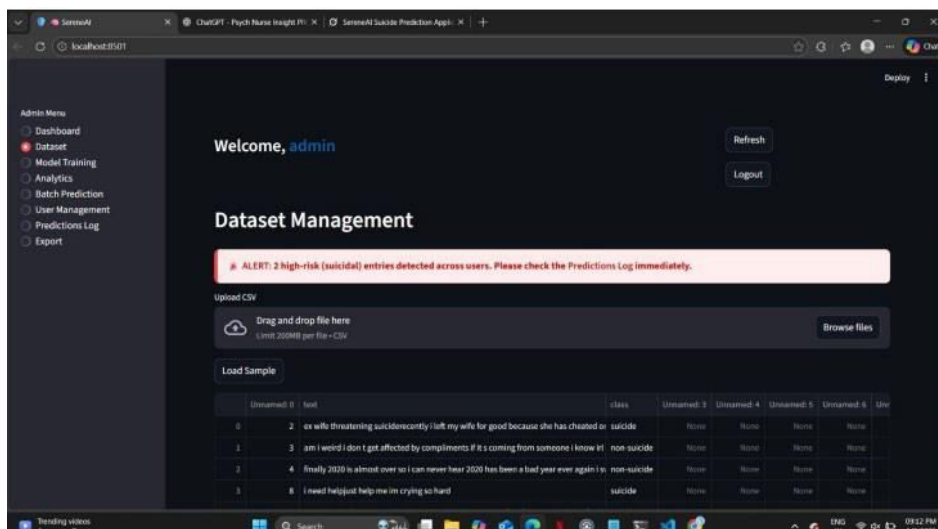


Figure 8: Admin View Dataset

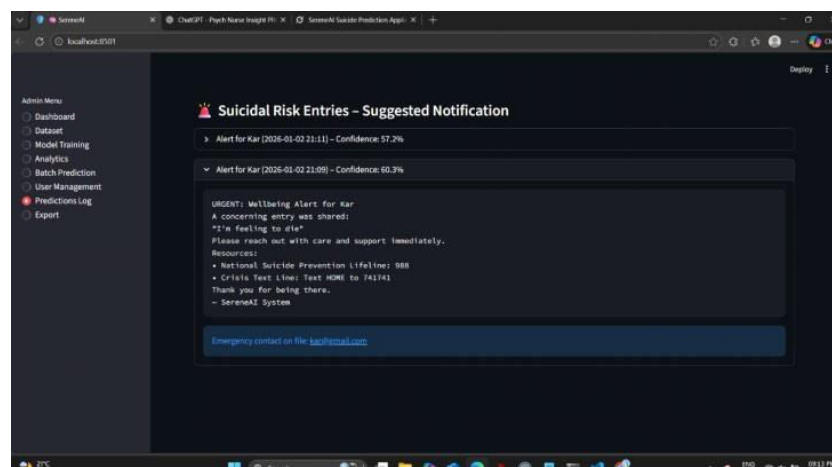


Figure 9: Admin View Prediction log

The admin dashboard successfully displayed real-time alerts for high-risk entries, triggered emergency notification workflows directing messages to registered emergency contacts, and provided a professional chatbot interface for immediate user support. Prediction logs were accessible for audit and model retraining purposes. These results confirm that the system meets its core mental health objectives: reducing detection latency, enabling proactive risk assessment, and providing decision-makers with actionable, interpretable insights through a privacy-preserving platform.

V. CONCLUSION

This paper presented SereneAI, a web-based platform designed to bridge the critical gap between the emergence of suicidal ideation and timely clinical intervention. By integrating NLP-based text normalization, supervised machine learning risk classification, role-based access control, and real-time alert mechanisms within a single lightweight platform, SereneAI directly addresses the limitations of existing paper-based and self-report mental health surveillance approaches.

The system demonstrated consistent performance across all tested scenarios, accurately classifying suicidal ideation risk and maintaining responsiveness under realistic usage conditions. Privacy-by-design principles ensured that sensitive user data remained protected, while the modular architecture supports straightforward extension to additional languages, data sources, or predictive features. Pilot deployment confirmed the system's effectiveness in generating timely alerts and connecting at-risk users to immediate professional support resources

As future work, SereneAI can be extended to incorporate real-time social media feed integration across multiple platforms, mobile application interfaces for broader accessibility, multilingual NLP support, advanced transformer-based prediction models, and automated integration with clinical electronic health record systems. These enhancements would position SereneAI as a comprehensive mental health intelligence platform suitable for deployment by healthcare providers, educational institutions, and non-governmental organizations operating in resource-constrained environments, potentially contributing to earlier intervention and the reduction of preventable loss of life.

REFERENCES

- [1] World Health Organization (WHO). (2022). Suicide worldwide statistics and prevention strategies. <https://www.who.int/>
- [2] American Psychological Association (APA). (2023). Mental health assessment guidelines. <https://www.apa.org/>
- [3] IEEE. (2022). Artificial intelligence applications in mental health monitoring. <https://ieeexplore.ieee.org/>
- [4] National Institute of Mental Health (NIMH). (2023). Suicide prevention and early detection research. <https://www.nimh.nih.gov/>
- [5] De Choudhury, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. Proceedings of ICWSM.
- [6] Coppersmith, A., Ngo, K., Leary, R., & Wood, A. (2016). Exploratory analysis of social media prior to suicide attempts. Computational Linguistics and Clinical Psychology Workshop.
- [7] Pestian, J., Nasrallah, H., Matykiewicz, P., Bennett, A., & Leenaars, A. (2012). Suicide note classification using NLP. Biomedical Informatics Insights.
- [8] Lin, Z., Lin, M., Liu, C., & Lane, T. (2020). Detecting suicide ideation in social media using deep learning. Proceedings of IEEE BIBM.
- [9] Pereira, F. T., de Oliveira, R. S. M., & da Silva, A. S. (2021). Predicting suicidal risk from social media posts using machine learning. Journal of Medical Internet Research.
- [10] Scikit-learn Documentation. (2023). Machine learning in Python. <https://scikit-learn.org/stable/>
- [11] Hugging Face. (2023). Transformers for natural language processing. <https://huggingface.co/docs>
- [12] Flask Documentation. (2023). Web framework for Python applications. <https://flask.palletsprojects.com/>



International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 9.971