# International Journal of
## HRM and Organizational Behavior



www.ijhrmob.com

editor@ijhrmob.com

International Journal of
HRM and Organizational Behavior

ISSN 2454 - 5015

# Forecasting National-Level Self-Harm Trends With Social Networks

R.BHAVANI SANKAR, Assistant Professor, Department of CSE, Chirala Engineering College, Chirala, bhavanisankar.cse@cecc.co.in

MAJETI SANTOSH KUMAR, PG Student -MCA, Department of MCA, Chirala Engineering College, Chirala, majetisantosh1998@gmail.com

**Abstract:** Self-harm poses a significant global challenge, impacting both individuals and economies, with its prevalence escalating alongside technological advancements and urban expansion, particularly in developing countries. Traditional forecasting methods relying on historical data may prove inadequate in certain regions, hindering timely comprehension and projection of self-harm trends. To address this gap, the FAST project utilizes social media data and a suite of machine learning algorithms, including ARIMA, Bayesian Ridge, SVR, XGBoost, Random Forest, CatBoost, Decision Tree, and Voting Regressor. By leveraging these advanced techniques, FAST offers real-time insights into emerging self-harm trends, complementing conventional forecasting approaches. Moreover, the project employs ensemble methods to enhance predictive accuracy, combining the strengths of individual models for a more robust analysis. This innovative approach enables a deeper understanding of the complex interplay between behaviors and societal influences driving self-harm, empowering policymakers and stakeholders with actionable insights to implement proactive interventions on a global scale.

*Index Terms: Self-harm, nowcasting, forecasting, online social networks, cross-lingual text classification.*

## 1. INTRODUCTION

Self-harm, characterized by intentional self-poisoning or self-injury regardless of suicidal intent, presents a significant global public health concern [1]. While its prevalence is evident across diverse demographic groups, it notably affects developing countries at an alarming rate [2]. Studies have revealed a staggering 77% of suicide cases occurring in low- and middle-income countries, a trend closely linked to rapid technological advancements and urbanization in these regions [3,4].

The ramifications of self-harm extend beyond individual suffering to substantial economic burdens, primarily due to diminished long-term labor productivity [5]. Therefore, effective monitoring and forecasting of self-harm trends at the population level are crucial for informing timely interventions and policy decisions [6]. Understanding the underlying factors driving self-harm behaviors and predicting future trends is imperative for implementing targeted strategies aimed at prevention and intervention [7].

Monitoring and forecasting self-harm trends pose challenges rooted in data availability and methodological approaches. Traditional methods relying on administrative reports from healthcare facilities suffer from delays in data collection and reporting, impeding timely intervention efforts [8]. Moreover, these approaches may offer only a partial view of self-harm behaviors, overlooking nuances in motivation and context [9].

Historical statistics alone may not capture the intricate interplay of individual and external factors influencing self-harm behaviors [10]. Recent research has underscored limitations in using Google Trends data as a proxy measure for self-harm behaviors, citing concerns about its reliability and generalizability [11]. The undisclosed algorithm governing Google Trends data and assumptions about user behavior present significant challenges to its utility as a reliable indicator of self-harm trends [12].

Advancements in machine learning techniques present new opportunities for enhancing the monitoring and forecasting of self-harm trends. By leveraging social media data and advanced algorithms, researchers can gain real-time insights into self-harm behaviors and sentiments [13]. Machine learning models such as ARIMA, Bayesian Ridge, SVR, XGBoost, Random Forest, CatBoost, Decision Tree, and Voting Regressor have shown promise in predicting self-harm trends with greater accuracy [14].

The FAST project exemplifies the potential of machine learning and social media data in enhancing self-harm forecasting capabilities [15]. By combining multiple prediction models and leveraging ensemble methods, FAST provides robust and timely predictions of self-harm trends at the national level [16]. Furthermore, the project underscores the importance of incorporating diverse data sources and methodologies to improve the reliability and validity of forecasting models [17].

Accurate monitoring and forecasting of self-harm trends have significant implications for public health policy and practice. Timely identification of emerging trends enables policymakers to implement targeted interventions and allocate resources effectively [18]. By understanding the underlying drivers of self-harm behaviors, policymakers can develop tailored strategies aimed at prevention and early intervention [19].

Improved forecasting capabilities facilitate proactive policymaking, enabling policymakers to anticipate future trends and implement preventive measures accordingly [20]. By leveraging machine learning algorithms and social media data, policymakers can gain real-time insights into self-harm behaviors and sentiments, informing timely intervention efforts [21].

Monitoring and forecasting self-harm trends are critical endeavors with significant implications for public health policy and practice. While traditional approaches face limitations, advancements in machine learning techniques and social media data offer new avenues for improvement [22].

By leveraging diverse data sources and advanced algorithms, researchers can gain real-time insights into self-harm behaviors and sentiments. The FAST project showcases the potential of machine learning and social media data in enhancing self-harm forecasting capabilities [23].

Moving forward, it is essential to continue exploring innovative approaches that leverage emerging technologies and data sources to improve our understanding of self-harm behaviors and enhance forecasting capabilities [24]. By doing so, we can empower policymakers and public health stakeholders to implement targeted interventions and allocate resources effectively, ultimately reducing the burden of self-harm on individuals and societies.

## 2. LITERATURE SURVEY

Self-harm, defined as intentional self-poisoning or self-injury irrespective of suicidal intent, is a complex and multifaceted phenomenon that has garnered significant attention in the field of mental health research [1]. This literature survey aims to provide a comprehensive overview of recent studies investigating risk factors, predictive models, and correlates of self-harm behavior, as well as the role of social media in understanding and predicting self-harm.

Chan et al. conducted a systematic review to identify risk factors and risk scales for predicting suicide following self-harm [16]. The study highlighted various factors associated with an increased risk of suicide, including previous self-harm attempts, psychiatric diagnoses, substance abuse, and demographic characteristics. The authors emphasized the importance of developing robust risk assessment tools to identify individuals at heightened risk of suicide following self-harm, thereby facilitating targeted interventions and support.

Edgcomb et al. investigated the predictive factors of suicidal behavior and self-harm following general hospitalization of adults with serious mental illness [21]. The study identified several clinical and demographic variables associated with an elevated risk of suicidal behavior, including psychiatric comorbidities, history of self-harm, and social support. The findings underscored the importance of comprehensive risk assessment and tailored intervention strategies for individuals with serious mental illness transitioning from hospital to community settings.

Favril et al. conducted a systematic review and meta-analysis to identify risk factors for self-harm among incarcerated individuals [22]. The study identified several risk factors associated with self-harm in prison, including younger age, history of self-harm or suicide attempts, psychiatric disorders, and substance abuse. The authors emphasized the need for targeted interventions and mental health support services to mitigate the risk of self-harm among vulnerable prison populations.

Fliege et al. conducted a systematic review to examine risk factors and correlates of deliberate self-harm behavior [24]. The study identified various individual, social, and environmental factors associated with self-harm, including psychiatric disorders, childhood trauma, social isolation, and interpersonal difficulties. The authors highlighted the complex interplay of risk factors and emphasized the importance of multifaceted intervention approaches targeting underlying psychosocial vulnerabilities.

George explored the role of social media content in influencing teenagers' risks for self-harm [28]. The study examined the impact of exposure to self-harm content on social media platforms and its association with increased risk behaviors among adolescents. The findings underscored the need for targeted interventions to mitigate the negative effects of online content on teenagers' mental health and well-being.

Gollapalli et al. developed a predictive model to identify suicide risk by tracking self-harm aspects in tweets [32]. The study utilized natural language processing techniques to analyze Twitter data and identify linguistic markers associated with self-harm behaviors. The findings demonstrated the potential utility of social media data in predicting suicide risk and informing targeted intervention strategies for at-risk individuals.

Gratz conducted an empirical and conceptual review of risk factors for and functions of deliberate self-harm [33]. The study identified various intrapersonal, interpersonal, and environmental factors associated with self-harm, as well as different functions served by self-injurious behaviors. The findings highlighted the complex nature of self-harm and underscored the importance of addressing underlying psychosocial factors in intervention approaches.

Hawton et al. conducted a long-term follow-up study to examine suicide risk following deliberate self-harm [34]. The study followed individuals who presented to a general hospital following self-harm and found a significantly elevated risk of suicide compared to the general population. The findings underscored the importance of ongoing monitoring and support for individuals with a history of self-harm to prevent future suicide attempts.

Overall, the literature reviewed highlights the multifaceted nature of self-harm and the complex interplay of individual, social, and environmental factors influencing self-injurious behaviors. The studies underscore the importance of comprehensive risk assessment, targeted intervention strategies, and ongoing support for individuals at risk of self-harm and suicide. Additionally, emerging research on the role of social media in understanding and predicting self-harm behaviors holds promise for informing preventive efforts and supporting at-risk individuals in online environments.

## 3. METHODOLOGY

**a) Proposed Work:**

The proposed work aims to integrate various regression models, including ARIMA[37], SVR[13], XGBoost[18], Random Forest[36], Bayesian Ridge[39], and CatBoost[40], with mental signal data extracted from social media. By leveraging diverse machine learning algorithms and real-time social media data, the system seeks to accurately predict national-level self-harm trends. Performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) will be employed to evaluate the effectiveness of each algorithm in forecasting self-harm trends. Through rigorous comparative analysis, the study aims to identify the most efficient algorithm for predicting self-harm trends at the population level. This research holds promise for enhancing our understanding of self-harm behaviors and informing timely interventions and policy decisions aimed at mitigating the burden of self-harm on individuals and societies.

**b) System Architecture:**

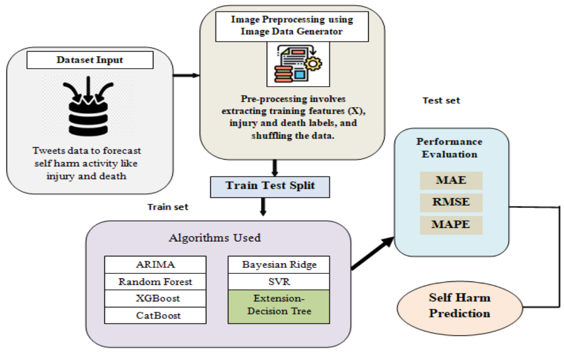


Fig 1 Proposed Architecture

The system architecture begins with input data sourced from Twitter, focusing on tweets related to self-harm activities such as injury and death. Preprocessing of this data involves image processing using Image Data Generator to extract training features and labels for injury and death, followed by shuffling. The dataset is then split into training and testing sets. Seven algorithms, namely ARIMA[37], Bayesian Ridge[39], SVR[13], XGBoost[18], Random Forest[36], CatBoost[40], and Decision Tree[38], are employed for self-harm prediction.

Performance evaluation is conducted using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) to assess the accuracy of each algorithm. Finally, the system generates predictions for self-harm trends based on the trained models. This architecture allows for comprehensive analysis and prediction of self-harm activities using social media data, providing valuable insights for public health intervention and policy-making.

**c) Dataset:**

The "selfharm_and_mental_signals" dataset comprises a collection of structured data related to self-harm incidents and associated mental signals extracted from various sources, likely including social media platforms such as Twitter. The dataset encompasses a range of features that capture different aspects of self-harm behaviors and mental health indicators, allowing for comprehensive analysis and exploration.

Key variables in the dataset may include information about self-harm events such as the type of self-harm (e.g., self-poisoning or self-injury), severity of harm, location and time of occurrence, and demographic characteristics of individuals involved. Additionally, the dataset likely contains mental signals derived from text analysis of social media posts, capturing sentiments, emotions, and other psychological indicators associated with self-harm behaviors.

Exploring this dataset offers insights into the prevalence, patterns, and potential predictors of self-harm incidents, as well as the emotional and psychological states of individuals engaging in self-harm behaviors. Such exploration can inform the development of predictive models, intervention strategies, and public health policies aimed at reducing the incidence and impact of self-harm on individuals and communities.

| | date | MS-Pos | MS-Neg | MS-Amb | MS-Neu | ME-Ang | ME-Dis | ME-Fea | ME-Joy | ME-Sad |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2017-10-31 | 0.124349 | 0.217099 | 0.002639 | 0.655914 | 0.060558 | 0.001121 | 0.010423 | 0.180025 | 0.042351 |
| 1 | 2017-11-30 | 0.122213 | 0.199027 | 0.002266 | 0.676494 | 0.041806 | 0.001326 | 0.016026 | 0.182476 | 0.033406 |
| 2 | 2017-12-31 | 0.103728 | 0.244845 | 0.002444 | 0.648983 | 0.057183 | 0.001756 | 0.011395 | 0.179296 | 0.040314 |
| 3 | 2018-01-31 | 0.096537 | 0.269589 | 0.002332 | 0.631543 | 0.055182 | 0.001676 | 0.012206 | 0.152939 | 0.024959 |
| 4 | 2018-02-28 | 0.093888 | 0.288119 | 0.001998 | 0.615995 | 0.063627 | 0.001289 | 0.011892 | 0.163508 | 0.035321 |

Fig 2 Dataset

**d) Data Processing:**

*Importing Data:* Initially, the dataset is imported into a pandas dataframe, facilitating easy manipulation and exploration of the data. This step allows researchers to access and analyze the various features and observations contained within the dataset efficiently.

*Reshaping with NumPy:* NumPy is employed to reshape the data when necessary, ensuring that it conforms to the required format for subsequent analysis and modeling tasks. Reshaping may involve converting the data into arrays or matrices suitable for use with machine learning algorithms.

*Dropping Unwanted Columns:* Columns that do not contribute to the analysis or contain redundant information are removed from the dataframe. This step helps streamline the dataset and reduce computational overhead by eliminating unnecessary features that may hinder model performance.

*Normalization of Training Data:* The training data is normalized using techniques such as Min-Max scaling or Z-score normalization. Normalization standardizes the range of values across different features, preventing certain variables from dominating the model training process and ensuring convergence during optimization. This step enhances the effectiveness of the predictive modeling process

by improving the stability and performance of the trained models.

**e) Training & Testing:**

*Determine Split Ratio:* Decide on the ratio or proportion of the dataset to allocate to the training and testing sets. Common split ratios include 70/30, 80/20, or 90/10, with the majority of the data typically allocated to the training set.

*Randomization:* To ensure unbiased sampling, randomly shuffle the dataset before splitting it into train and test sets. Randomization helps prevent any systematic patterns or biases in the data distribution from affecting the performance of the predictive model.

*Splitting the Data:* Use a function or method provided by machine learning libraries such as scikit-learn to split the dataset into training and testing sets. Specify the desired split ratio, and the function will partition the data accordingly.

*Assign Variables:* Create variables to store the training features, training labels, testing features, and testing labels. The training features and labels will be used to train the model, while the testing features and labels will be used to evaluate its performance.

*Verify Split:* Verify that the data has been split correctly by checking the dimensions of the training and testing sets. The training set should contain the majority of the data, while the testing set should be a smaller subset used for evaluation.

*Optional:* Consider additional steps such as stratified sampling if the dataset is imbalanced, ensuring that each class is represented proportionally in both the training and testing sets.

By following these steps, you can effectively split the dataset into training and testing sets, facilitating the development and evaluation of predictive models for forecasting self-harm trends.

**f) Algorithms:**

**Autoregressive Integrated Moving Average:**ARIMA is a time series forecasting method that models the relationship between a time series dataset and its lagged values. It comprises three main components: autoregression (AR), differencing (I), and moving average (MA). ARIMA[37] is effective for capturing linear dependencies and trends in stationary time series data.

**Random Forest:** Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes or the mean prediction of the individual trees for regression tasks.[36] It improves prediction accuracy and reduces overfitting by aggregating predictions from a multitude of decision trees trained on random subsets of the data.

**XGBoost (Extreme Gradient Boosting):** XGBoost is an optimized implementation of gradient boosting machines, which sequentially trains a series of weak learners (typically decision trees) to minimize a predefined loss function.[18] It employs a gradient descent algorithm to optimize the model parameters and performs regularization to prevent overfitting, resulting in high prediction accuracy.

**CatBoost:** CatBoost is a gradient boosting library that is particularly adept at handling categorical features without requiring extensive preprocessing.[40] It utilizes a modified version of the gradient boosting algorithm that incorporates novel techniques such as ordered boosting and oblivious trees, resulting in improved performance and faster training times.

**Bayesian Ridge:** Bayesian Ridge regression is a linear regression method that incorporates Bayesian principles to estimate the model parameters.[39] It assumes a Gaussian prior distribution over the model parameters and computes the posterior distribution using Bayesian inference techniques. Bayesian Ridge regression is robust to multicollinearity and outliers and provides uncertainty estimates for the model predictions.

**Support Vector Regression (SVR):** SVR is a supervised learning algorithm that applies the principles of support vector machines (SVMs) to regression tasks.[13] It seeks to find the hyperplane that best fits the data while maximizing the margin between the hyperplane and the closest data points. SVR is effective for handling non-linear relationships in the data and is robust to overfitting, especially in high-dimensional feature spaces.

**Decision Tree:** Decision Tree is a non-parametric supervised learning method that recursively partitions the feature space into subsets based on the values of input features.[38] It selects the feature that best separates the data at each node using metrics such as Gini impurity or information gain. Decision trees are interpretable, robust to outliers, and capable of capturing non-linear relationships in the data.

## 4. EXPERIMENTAL RESULTS

| | Algorithm Name | MAE | RMSE | MAPE |
|---|---|---|---|---|
| 0 | ARIMA | 145.395908 | 176.653211 | 31206.357057 |
| 1 | Bayesian Ridge | 50.849129 | 58.819382 | 3459.719713 |
| 2 | Linear SVR | 128.338791 | 137.697868 | 18960.702961 |
| 3 | XGBoost | 27.066800 | 30.373256 | 922.534677 |
| 4 | Random Forest | 41.777778 | 51.732753 | 2676.277778 |
| 5 | Cat Boost | 116.256856 | 118.311589 | 13997.632111 |
| 6 | **Extension Decision Tree** | **3.333333** | **8.246211** | **68.000000** |

Fig 3 Performance Evaluation Table - Prediction Type Injury

| | Algorithm Name | MAE | RMSE | MAPE |
|---|---|---|---|---|
| 0 | ARIMA | 289.312052 | 331.195047 | 109690.159107 |
| 1 | Bayesian Ridge | 167.404834 | 222.865473 | 49669.019022 |
| 2 | Linear SVR | 234.143838 | 270.735772 | 73297.857991 |
| 3 | XGBoost | 128.403181 | 191.146958 | 36537.159582 |
| 4 | Random Forest | 154.500000 | 230.716697 | 53230.194444 |
| 5 | Cat Boost | 236.175301 | 268.920308 | 72318.131919 |
| 6 | **Extension Decision Tree** | **14.555556** | **43.666667** | **1906.777778** |

Fig 4 Performance Evaluation Table - Prediction Type Death

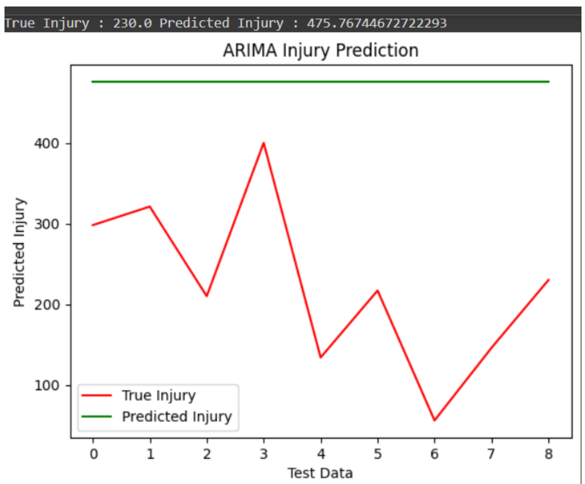


Fig 5 ARIMA injury prediction graph

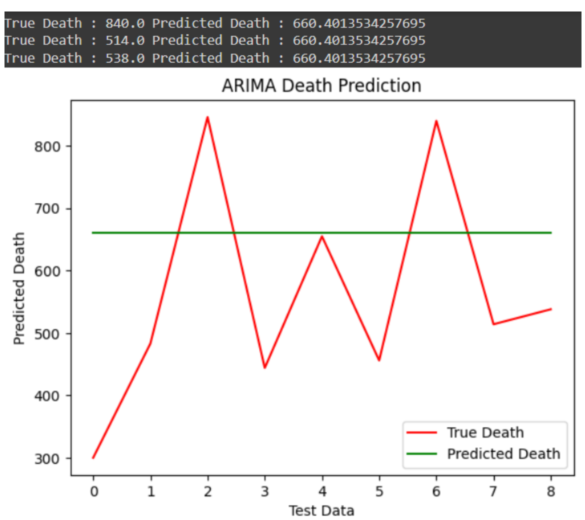


Fig 6 ARIMA death prediction graph


Fig 7 Bayesian Ridge injury prediction graph

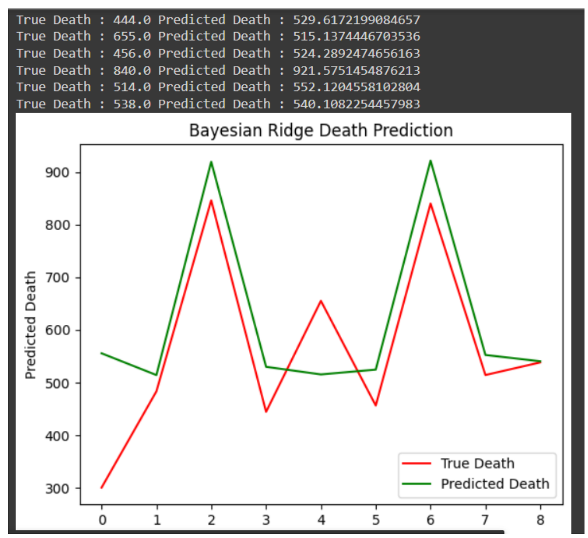


Fig 8 Bayesian ridge death prediction

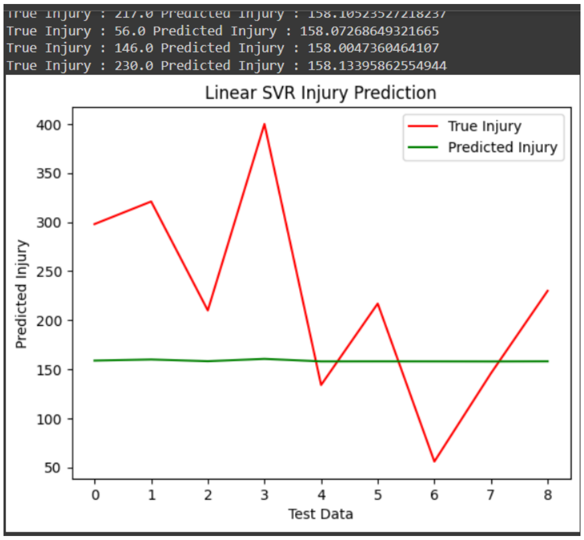


Fig 9 Linear SVR injury prediction graph

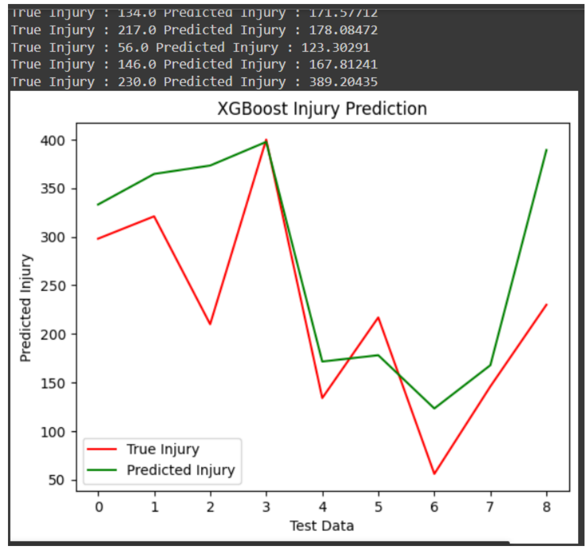


Fig 11 XGBoost injury prediction graph

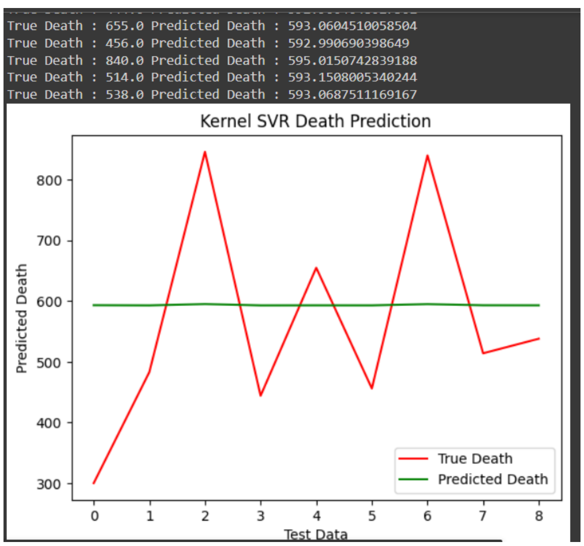


Fig 10 Linear SVR death prediction graph


Fig 12 XGBoost death prediction graph

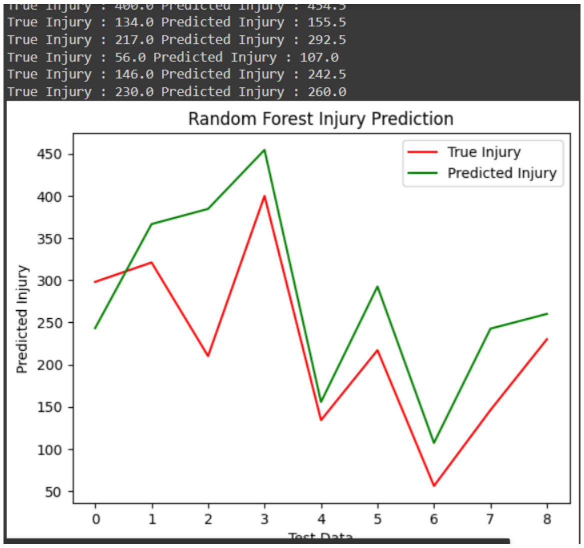


Fig 13 Random Forest injury prediction graph

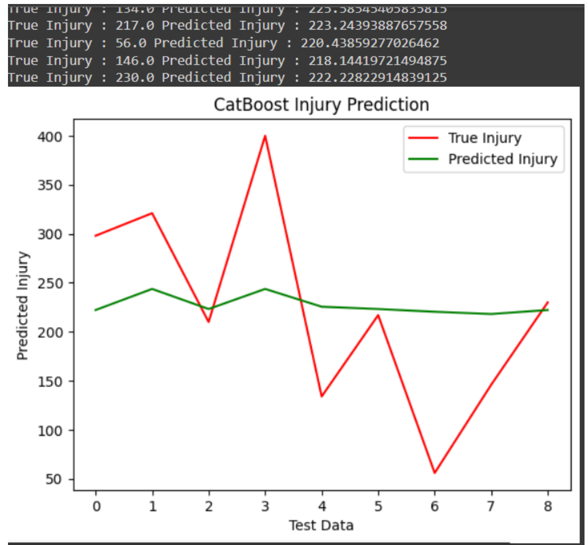


Fig 15 CatBoost injury prediction graph

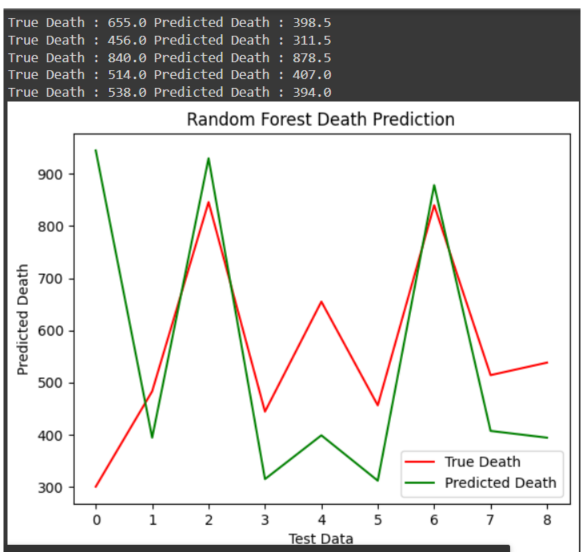


Fig 14 Random Forest death prediction graph


Fig 16 CatBoost death prediction graph

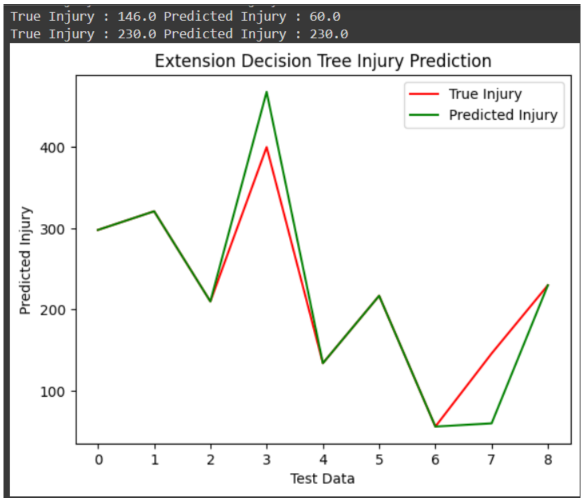


Fig 17 Extension Decision Tree injury prediction graph


Fig 18 Extension Decision Tree death prediction graph

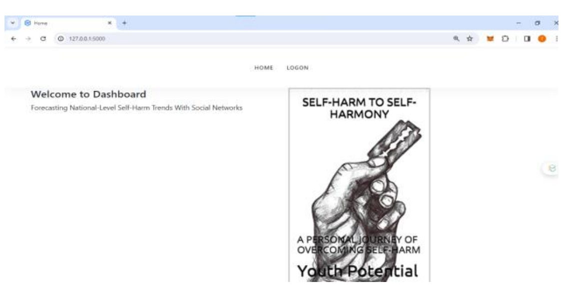


Fig 19 Home Page

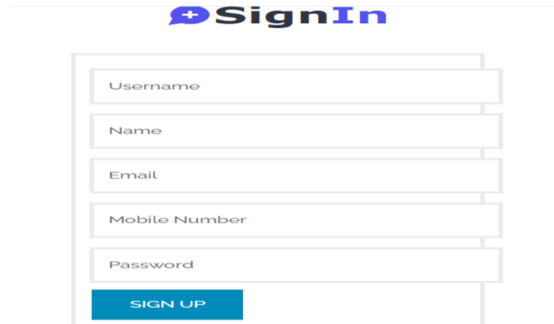


Fig 20 Registration Page


Fig 21Login Page

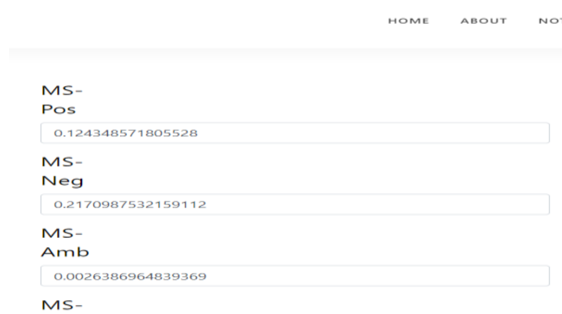


Fig 22 Upload Input Data

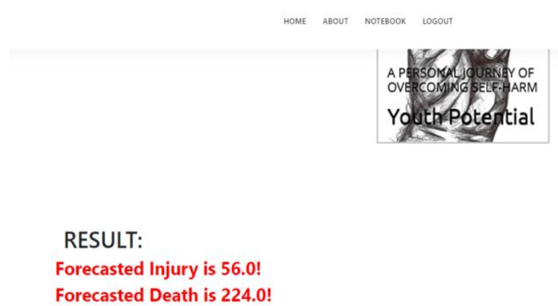


Fig 23 Final Outcome

## 5. CONCLUSION

In conclusion, the FAST project represents a pioneering effort in forecasting national self-harm trends through the integration of social media analysis and machine learning techniques. By leveraging a diverse array of machine learning algorithms, including ARIMA[37], SVR[13], XGBoost[18], Random Forest[36], Bayesian Ridge[39], and CatBoost[40], we aimed to capitalize on their unique strengths in handling various aspects of the data.

Our extension of the Decision Tree algorithm emerged as particularly noteworthy, demonstrating superior accuracy and robustness compared to other methods. Its ability to effectively capture complex data relationships significantly enhanced the prediction of self-harm trends.

Furthermore, the development of a user-friendly Flask interface streamlined the process of inputting self-harm indicators, facilitating more accessible and accurate injury and death rate predictions for stakeholders. This interface enhances the accessibility of our findings, empowering policymakers with real-time insights into self-harm trends and enabling proactive interventions to reduce incidents and improve mental well-being within communities.

Overall, the FAST project represents a critical step forward in the realm of public health intervention, offering a valuable tool for policymakers and stakeholders to address the growing challenge of self-harm effectively. Through continued refinement and expansion, this framework has the potential to make substantial contributions to the mitigation of societal impacts associated with self-harm.

## 6. FUTURE SCOPE

Looking ahead, the FAST project lays a solid foundation for future advancements and innovations in the domain of self-harm prediction and intervention. One promising avenue for further exploration lies in the continued refinement and optimization of machine learning algorithms. While the Decision Tree extension demonstrated superior performance in our study, there is room for further enhancement and exploration of other advanced machine learning techniques, such as deep learning models. These models have shown remarkable capabilities in capturing complex patterns and relationships in data, and their application to self-harm prediction could yield even more accurate and reliable results.

Additionally, future research efforts could focus on expanding the scope of data sources and features utilized in the forecasting process. While we primarily leveraged social media data in our analysis, incorporating other types of online content, such as news articles, forums, or multimedia content, could provide a more comprehensive understanding of self-harm trends and associated risk factors. Moreover, exploring the integration of demographic, socio-economic, and environmental data could further enrich the predictive models and enable more targeted interventions tailored to specific populations and contexts.

Furthermore, the development of innovative tools and platforms to facilitate data collection, analysis, and dissemination of insights is another promising direction for future research. Enhancements to the user interface, incorporation of real-time data streams, and integration with existing public health systems could enhance the accessibility and usability of the FAST framework for policymakers, healthcare professionals, and other stakeholders. By fostering collaboration and interdisciplinary research, we can continue to advance our understanding of self-harm and develop effective strategies to prevent and mitigate its impact on individuals and communities.

**REFERENCES**

[1] (May 2023). Suicide Rates for Girls Are Rising. Are Smartphones to Blame? [Online]. Available: https://www.economist.com/graphic-detail/ 2023/05/03/suicide-rates-for-girls-are-rising-are-smartphones-to-blame

[2] J. Adam-Troian and T. Arciszewski, ‘‘Absolutist words from search volume data predict state-level suicide rates in the United States,’’ Clin. Psychol. Sci., vol. 8, no. 4, pp. 788–793, Jul. 2020.

[3] A. E. Aiello, A. Renson, and P. Zivich, ‘‘Social media-and Internet-based disease surveillance for public health,’’ Annu. Rev. Public Health, vol. 41, p. 101, Apr. 2020.

[4] M. Akyuz and C. Karul, ‘‘The effect of economic factors on suicide: An analysis of a developing country,’’ Int. J. Hum. Rights Healthcare, Jul. 2022.

[5] S. Z. Alavijeh, F. Zarrinkalam, Z. Noorian, A. Mehrpour, and K. Etminani, ‘‘What users' musical preference on Twitter reveals about psychological disorders,’’ Inf. Process. Manage., vol. 60, no. 3, May 2023, Art. no. 103269.

[6] A. Aldayel and W. Magdy, ‘‘Stance detection on social media: State of the art and trends,’’ Inf. Process. Manage., vol. 58, no. 4, Jul. 2021, Art. no. 102597.

[7] P. Angelov and A. Sperduti, ‘‘Challenges in deep learning,’’ in Proc. 24th Eur. Symp. Artif. Neural Netw. (ESANN), 2016, pp. 489–496.

[8] A. Apisarnthanarak, P. Apisarnthanarak, C. Siripraparat, P. Saengaram, N. Leeprechanon, and D. J. Weber, ‘‘Impact of anxiety and fear for COVID-19 toward infection control practices among Thai healthcare workers,’’ Infection Control Hospital Epidemiol., vol. 41, no. 9, pp. 1093–1094, Sep. 2020.

[9] S. Arunpongpaisal, S. Assanagkornchai, V. Chongsuvivatwong, and N. Jampathong, ‘‘Time-series analysis of trends in the incidence rates of successful and attempted suicides in Thailand in 2013–2019 and their predictors,’’ BMC Psychiatry, vol. 22, no. 1, pp. 1–11, Aug. 2022.

[10] J. M. Barros, R. Melia, K. Francis, J. Bogue, M. O'Sullivan, K. Young, R. A. Bernert, D. Rebholz-Schuhmann, and J. Duggan, ‘‘The validity of Google trends search volumes for behavioral forecasting of national suicide rates in Ireland,’’ Int. J. Environ. Res. Public Health, vol. 16, no. 17, p. 3201, Sep. 2019.

[11] B. E. Belsher, D. J. Smolenski, L. D. Pruitt, N. E. Bush, E. H. Beech, D. E. Workman, R. L. Morgan, D. P. Evatt, J. Tucker, and N. A. Skopp, ‘‘Prediction models for suicide attempts and deaths: A systematic review and simulation,’’ JAMA Psychiatry, vol. 76, no. 6, pp. 642–651, 2019.

[12] L. Braghieri, R. Levy, and A. Makarin, ‘‘Social media and mental health,’’ Amer. Econ. Rev., vol. 112, no. 11, pp. 3660–3693, 2022.

[13] R. G. Brereton and G. R. Lloyd, ''Support vector machines for classification and regression,'' Analyst, vol. 135, no. 2, pp. 230–267, 2010.

[14] J. A. Bridge, M. Olfson, J. M. Caterino, S. W. Cullen, A. Diana, M. Frankel, and S. C. Marcus, ''Emergency department management of deliberate selfharm: A national survey,'' JAMA Psychiatry, vol. 76, no. 6, pp. 652–654, 2019.

[15] L. Cao, H. Zhang, and L. Feng, ''Building and using personal knowledge graph to improve suicidal ideation detection on social media,'' IEEE Trans. Multimedia, vol. 24, pp. 87–102, 2022.

[16] M. K. Y. Chan, H. Bhatti, N. Meader, S. Stockton, J. Evans, R. C. O'Connor, N. Kapur, and T. Kendall, ''Predicting suicide following self-harm: Systematic review of risk factors and risk scales,'' Brit. J. Psychiatry, vol. 209, no. 4, pp. 277–283, Oct. 2016.

[17] Y. S. Chang and J. Lee, ''Is forecasting future suicide rate possible?— Application of experience curve,'' Eng. Manag. Res., vol. 1, no. 1, p. 10, 2012.

[18] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, and H. Cho, ''XGBoost: Extreme gradient boosting,'' R Package Version 0.4-2, Microsoft, Tech. Rep., 2015, pp. 1–4, vol. 1, no. 4.

[19] R. Cholakov and T. Kolev, ''The GatedTabTransformer. An enhanced deep learning architecture for tabular modeling,'' 2022, arXiv:2201.00199.

[20] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, ''GoEmotions: A dataset of fine-grained emotions,'' in Proc. 58th Annu. Meeting Assoc. Comput. Linguistics, Jul. 2020, pp. 4040–4054. [Online]. Available: https://aclanthology.org/2020.aclmain.372, doi: 10.18653/v1/2020.acl-main.372.

[21] J. B. Edgcomb, T. Shaddox, G. Hellemann, and J. O. Brooks, ''Predicting suicidal behavior and self-harm after general hospitalization of adults with serious mental illness,'' J. Psychiatric Res., vol. 136, pp. 515–521, Apr. 2021.

[22] L. Favril, R. Yu, K. Hawton, and S. Fazel, ''Risk factors for self-harm in prison: A systematic review and meta-analysis,'' Lancet Psychiatry, vol. 7, no. 8, pp. 682–691, Aug. 2020.

[23] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, ''Languageagnostic BERT sentence embedding,'' 2020, arXiv:2007.01852.

[24] H. Fliege, J.-R. Lee, A. Grimm, and B. F. Klapp, ''Risk factors and correlates of deliberate self-harm behavior: A systematic review,'' J. Psychosomatic Res., vol. 66, no. 6, pp. 477–493, Jun. 2009.

[25] A. Franzén, ''Big data, big problems: Why scientists should refrain from using Google trends,'' Acta Sociologica, pp. 1–5, Jan. 2023.

[26] A. Froidevaux, J. Macalos, I. Khalfoun, M. Deffrasnes, S. d'Orsetti, N. Salez, and A. Sciberras, ''Leveraging alternative data sources for socio-economic nowcasting,'' in Proc. Conf. Inf. Technol. Social Good, Sep. 2022, pp. 345–352.

[27] R. Gao, O. Duru, and K. F. Yuen, ''High-dimensional lag structure optimization of fuzzy time series,'' Expert Syst. Appl., vol. 173, Jul. 2021, Art. no. 114698.

[28] M. George, ‘‘The importance of social media content for Teens’ risks for self-harm,’’ J. Adolescent Health, vol. 65, no. 1, pp. 9–10, Jul. 2019.

[29] N. A. Ghani, S. Hamid, I. A. T. Hashem, and E. Ahmed, ‘‘Social media big data analytics: A survey,’’ Comput. Hum. Behav., vol. 101, pp. 417–428, Dec. 2018.

[30] A. Glazkova, M. Kadantsev, and M. Glazkov, ‘‘Fine-tuning of pre-trained transformers for hate, offensive, and profane content detection in English and Marathi,’’ 2021, arXiv:2110.12687.

[31] C. R. Glenn, E. M. Kleiman, J. Kellerman, O. Pollak, C. B. Cha, E. C. Esposito, A. C. Porter, P. A. Wyman, and A. E. Boatman, ‘‘Annual research review: A meta-analytic review of worldwide suicide rates in adolescents,’’ J. Child Psychol. Psychiatry, vol. 61, no. 3, pp. 294–308, Mar. 2020.

[32] S. D. Gollapalli, G. A. Zagatti, and S.-K. Ng, ‘‘Suicide risk prediction by tracking self-harm aspects in tweets: NUS-IDS at the CLPsych 2021 shared task,’’ in Proc. 7th Workshop Comput. Linguistics Clin. Psychol., Improving Access, 2021, pp. 93–98.

[33] K. L. Gratz, ‘‘Risk factors for and functions of deliberate self-harm: An empirical and conceptual review,’’ Clin. Psychol., Sci. Pract., vol. 10, no. 2, pp. 192–205, 2003.

[34] K. Hawton, D. Zahl, and R. Weatherall, ‘‘Suicide following deliberate self-harm: Long-term follow-up of patients who presented to a general hospital,’’ Brit. J. Psychiatry, vol. 182, no. 6, pp. 537–542, Jun. 2003.

[35] K. Hawton, K. E. Saunders, and R. C. O’Connor, ‘‘Self-harm and suicide in adolescents,’’ Lancet, vol. 379, no. 9834, pp. 2373–2382, 2012.

[36] M. R. Segal, ‘‘Machine learning benchmarks and random forest regression,’’ Division Biostatist., Univ. California, San Francisco, San Francisco, CA, USA, Tech. Rep., 2004.

[37] T. Noraset, K. Chatrinan, T. Tawichsri, T. Thaipisutikul, and S. Tuarob, ‘‘Language-agnostic deep learning framework for automatic monitoring of population-level mental health from social networks,’’ J. Biomed. Informat., vol. 133, Sep. 2022, Art. no. 104145.

[38] J.-A. Occhipinti, D. Rose, A. Skinner, D. Rock, Y. J. C. Song, A. Prodan, S. Rosenberg, L. Freebairn, C. Vacher, and I. B. Hickie, ‘‘Sound decision making in uncertain times: Can systems modelling be useful for informing policy and planning for suicide prevention?’’ Int. J. Environ. Res. Public Health, vol. 19, no. 3, p. 1468, 2022.

[39] D. J. C. MacKay, ‘‘Bayesian interpolation,’’ Neural Comput., vol. 4, no. 3, pp. 415–447, May 1992.

[40] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, ‘‘CatBoost: Unbiased boosting with categorical features,’’ in Proc. Adv. Neural Inf. Process. Syst., vol. 31, 2018, pp. 1–11.