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

Forecasting National-Level Self-Harm Trends with Social Networks using Machine Learning Algorithms

N.Ramadevi[1], Paradesi Subba Rao[2], A. Satya Nymisha[3], K. Pallavi[4]
M. Jeba Samreen[5], E. Meghana[6], S. Ruksana Parveen[7]

[1]Associate Professor, Department of Computer Science and Engineering (Data Science), Santhiram Engineering College, Nandyal

[2]Assistant Professor, Department of Computer Science and Engineering , Santhiram Engineering College, Nandyal

[3, 4,5,6,7] Department of Computer Science and Engineering (Data Science), Santhiram Engineering College, Nandyal

E-mail: ramadevi.cse@srecnandyal.edu.in

Abstract: Self-harm pertains to actions of self-inflicted poisoning or injury that lead to either nonfatal injuries or death, irrespective of the individual's intention. Self-harm incidents not only cause loss to individuals but also incur a negative impact on the nation's economy. Studies have demonstrated an increase in trends of self-harm that are correlated with the emergence of technological advancements and swift urban expansion in developing countries. The capacity to nowcast and forecast national-level patterns of self-harm trends could be imperative to policymakers and stakeholders in the public health sector, as it would enable them to implement prompt measures to counteract the underlying factors or avert these projected calamities. Prior research has utilized historical data to predict self-harm trends at the population level in various nations using conventional statistical forecasting methods. However, in some countries, such historical statistics may be challenging to obtain or insufficient for accurate prediction, impeding the ability to comprehend and project the national self-harm landscape in a timely manner. This paper proposes FAST, a framework designed to forecast self-harm patterns at the national level by analyzing mental signals obtained from a large volume of social media data. These signals serve as a proxy for real-world population mental health that could be used to enhance the forecastability of self-harm trends. Specifically, language-agnostic language models are first trained to extract different mental signals from collected social media messages. Then, these signals are aggregated and processed into multi-variate time series, on which the time-delay embedding algorithm is applied to transform into temporal embedded instances. Finally, various machine learning regressors are validated for their forecastability. The proposed method is validated through a case study in Thailand, which utilizes a set of 12 mental signals extracted from tweets to forecast death and injury cases resulting from selfharm. The results show that the proposed method outperformed the traditional ARIMA baseline by 43.56% and 36.48% on average in terms of MAPE on forecasting death and injury cases from self-harm, respectively. As far as current understanding permits, our research represents the initial exploration of utilizing aggregated social media information for the purposes of nowcasting and forecasting trends of self-harm on a nationwide scale. The results not only provide insight into improved forecasting techniques for self-harm trends but also establish a foundation for forthcoming social-network-driven applications that hinge on the

capacity to predict socioeconomic factors. We further experimented with Decision Tree algorithm & Voting regressor which is the best algorithms in machine learning and this algorithms are giving lesser MAE error compare to other algorithms.

Key Words - Self-harm, nowcasting, forecasting, online social networks, cross-lingual text classification.

1. INTRODUCTION

Self-harm refers to intentional self-poisoning or self-injury, regardless of the nature of the motivation or the severity of suicidal intent, that could result in injury or death. Self-harm and suicide have been prevalent problems, especially in developing countries. According to a recent study, a significant proportion of suicide cases, approximately 77%, were observed in low- and middle-income countries. This trend has been associated with the uptake of technological advancements and the rapid pace of urbanization in these regions. The exacerbation of incidents involving self-harm not only results in personal grief and loss but also has enduring adverse effects on the economy, primarily due to the reduction in long-term labor productivity. The ability to monitor and forecast population-level self-harm trends could prove vital to national-level policymakers and public healthcare stakeholders in devising means to timely gauge the situations and implement procedures to neutralize or prevent such anticipated tragedies. For example, upon being informed that certain stringent policies aimed at addressing nationwide epidemics have led to mental health issues among citizens and are expected to contribute to a significant increase in self-harm trends, policymakers may consider making appropriate modifications to the current policies that are causing these issues. Furthermore, the implementation of public health interventions, such as mobile psychiatry units or hotlines, could be deployed to target populations experiencing adverse effects. Presently, the techniques employed to acquire knowledge about selfharm trends at the national level depend on administrative reports from healthcare centers and hospitals across the country. This approach necessitates significant financial, human, and time resources, leading to infrequent and delayed data availability. Statistics that are coarse-grained and delayed may have limited utility for the purpose of proactive policy-making.

This paper introduces FAST, a framework for forecasting national self-harm patterns using mental signals extracted from social media data. Language-agnostic models analyze messages, creating multivariate time series transformed by time-delay embedding. Machine learning regressors, including Decision Tree and Voting, outperform traditional methods in forecasting death and injury cases from self-harm, offering valuable insights for policymakers.

The rise in self-harm incidents, linked to technological advancements and urban expansion in developing nations, poses significant challenges for timely prediction and understanding at the national level. Conventional methods relying on historical data may be insufficient. This study proposes FAST, utilizing social media-derived mental signals to forecast self-harm trends, addressing the need for more effective predictive models in public health.

2. LITERATURE SURVEY

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

[37] This study delves into the intricate relationship between bullying victimization, self-harm, and suicidality, examining gender differences in the mediating effects of depression and anxiety. Analyzing data from 2522 Australian adolescents aged 12–17, the research, based on Baron and Kenny's approach, investigates the nuanced impact on boys and girls. Out of the 31.1% bullied victims, 53.2% were girls, and 46.8% were boys, revealing a higher prevalence of depression, anxiety, self-harm, and suicidality among girls. Logistic regressions and the Sobel test reveal that depression significantly mediates the association between bullying victimization and self-harm and suicidality in both genders. However, anxiety disorder serves as a mediating factor exclusively in girls. These results emphasize the pronounced impact of bullying on girls, highlighting the need for gender-specific prevention programs. The study's cross-sectional design and reliance on self-reported data introduce potential limitations, cautioning against inferring causality. Overall, the findings underscore the urgency of tailored interventions to address the differential psychological consequences of bullying on boys and girls, emphasizing the role of depression and anxiety in mediating these complex associations.

[21] In this longitudinal cohort study, focusing on 15,644 general non-psychiatric hospitalizations of adults with serious mental illnesses such as depression, bipolar, and psychotic disorders, the research aims to predict the risk of readmission for suicidal behavior and self-harm following general hospitalization. Utilizing structured electronic health record data from an urban health system in the southwestern United States between 2006 and 2017, supervised machine learning, specifically the Classification and Regression Tree algorithm, was applied to predict the risk of suicide attempt and self-harm in the subsequent year based on data from one year prior to and including the index hospitalization. The algorithm yielded a high classification prediction with an area under the receiver operating curve (AUC) of 0.86, indicating robust predictive capabilities. The study discerned varying incidence rates of suicide-related behavior, with the highest observed after hospitalizations of individuals with prior suicide attempts or self-harm. Predictor combinations, particularly concomitant alcohol use disorder with moderate medical morbidity and age ≤55 with low medical morbidity, played a pivotal role in explaining the majority of the identified risks. These findings highlight the potential of efficient and interpretable machine learning algorithms in informing clinical decision support, resource allocation, and preventive interventions for adults with serious mental illnesses undergoing general hospitalization.

[32] The study presents a system developed for the CLPsych 2021 Shared Task, focusing on identifying users at risk of suicide based on their tweets. Grounded in mental health research linking self-harm tendencies with suicide, the system employs the Self-Harm Topic Model (SHTM). This model combines Latent Dirichlet Allocation with a self-harm dictionary to characterize self-harm aspects expressed in users' tweets over a specified time frame. The incorporation of SHTM allows for a nuanced analysis of daily tweets, capturing variations in moods and topics over time. Subsequently, these differences serve as features to train a deep learning model for suicide prediction. By integrating both topic modeling and deep learning approaches, the system aims to provide a comprehensive understanding of users' expressions related to self-harm, contributing

to a more accurate prediction of suicide risk. The research underscores the potential of leveraging social media data and advanced modeling techniques for proactive suicide risk identification, emphasizing the importance of computational methods in augmenting mental health surveillance and intervention efforts.

[22] This systematic review and meta-analysis investigate risk factors for self-harm in prison populations, addressing a significant concern for prisoner morbidity. The study, encompassing 35 independent studies from 20 countries and involving 663,735 prisoners, identifies suicide-related antecedents as the most robust predictors of self-harm in prison. Current or recent suicidal ideation, a lifetime history of suicidal ideation, and previous self-harm exhibit strong associations with self-harm, emphasizing the critical role of mental health factors. Additionally, any current psychiatric diagnosis, major depression, and borderline personality disorder significantly contribute to self-harm risk. Prison-specific environmental factors, including solitary confinement, disciplinary infractions, and experiences of sexual or physical victimization while in prison, also emerged as notable contributors. In contrast, sociodemographic and criminological factors displayed modest associations with self-harm in prison. The comprehensive range of identified risk factors underscores the necessity for a holistic, prison-wide approach to self-harm prevention. The study advocates for multiagency collaboration between mental health, social care, and criminal justice services to implement population-wide and targeted strategies effectively, addressing the complex interplay of clinical and custody-related domains in mitigating self-harm risk within prison settings.

[36] This study, utilizing data from the COVID-19 Social Study with a sample size of 44,775 individuals during the initial month of the COVID-19 pandemic, delves into the prevalence of abuse, self-harm, and thoughts of suicide/self-harm in the UK. The findings reveal heightened reported frequencies of these distressing experiences, particularly among women, individuals from Black, Asian, and minority ethnic (BAME) groups, and those facing socioeconomic challenges, unemployment, disability, chronic physical illnesses, mental disorders, and a COVID-19 diagnosis. Notably, psychiatric medications emerged as the most commonly used form of support, yet less than half of the affected individuals were accessing formal or informal support systems. These patterns underscore the multifaceted impact of the pandemic, reflecting disparities across demographic and socioeconomic lines. The study signals a need for targeted interventions and support structures, especially for vulnerable populations, to address the increased risks of abuse, self-harm, and suicidal ideation during times of societal crisis. Additionally, the observed underutilization of available support underscores the importance of enhancing accessibility and awareness of mental health resources to mitigate the adverse effects of the pandemic on individuals' mental well-being.

3. METHODOLOGY

i) Proposed Work:

This paper proposes FAST, a framework designed to forecast self-harm patterns at the national level by analyzing mental signals obtained from a large volume of social media data. These signals serve as a proxy for real-world population mental health that could be used to enhance the forecastability of self-

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

harm trends. Specifically, language-agnostic language models are first trained to extract different mental signals from collected social media messages. Then, these signals are aggregated and processed into multi-variate time series, on which the time-delay embedding algorithm is applied to transform into temporal embedded instances. Finally, various machine learning regressors are validated for their forecastability. One of the greatest machine learning algorithms, the Decision Tree algorithm & Voting regressor, what we have experimented. Compared to other algorithms, it produces a lower MAE error.

ii) System Architecture:

The forecasting system for national-level self-harm trends with social networks comprises several key components. The initial phase involves exploring the "selfharm_and_mental_signals" dataset, focusing on relevant features. Following this, data processing techniques are applied to clean and prepare the dataset for analysis.

The dataset is then split into training and testing sets to facilitate model evaluation. The forecasting models include ARIMA, Bayesian Ridge, Support Vector Regression (SVR), XGBoost, Random Forest, CatBoost, Decision Tree, and a Voting Regressor ensemble.

The training and building of each model are executed using the respective algorithms, with hyperparameter tuning where applicable. Performance evaluation metrics such as accuracy, precision, recall, and F1 score are employed to assess the effectiveness of each model in predicting self-harm trends.

The system architecture emphasizes a modular approach, incorporating diverse algorithms to enhance prediction accuracy. The Voting Regressor combines the strengths of individual models. The architecture ensures a comprehensive analysis of national-level self-harm trends, leveraging social network data for improved forecasting precision.

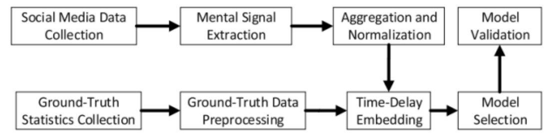

Fig 1: System Architecture

iii) Dataset Collection:

The dataset for this research is integral to exploring the relationship between mental signals extracted from social media data and the forecasting of self-harm cases at the national level. The study focuses on Thailand, employing a case study approach that integrates social media data with ground-truth statistics of self-harm cases.

Two distinct datasets are utilized in this research. The first dataset comprises approximately 4.9 million tweets collected randomly using the Twitter API from October 2017 to January 2021. To ensure generalizability across various social media platforms, only the timestamp and textual information from each tweet are retained. This language-agnostic approach in extracting mental signals allows the proposed forecasting framework to be applicable to multivariate time series data, fostering generalization across linguistic and geographical contexts.

The second dataset consists of historical statistics of monthly cases of death and injury from self-harm in Thailand. This dataset is sourced from the Department of Mental Health, Ministry of Public

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

Health of Thailand, providing ground-truth information for model validation. The dataset includes monthly numbers of tweets (displayed as a bar chart with the right Y-axis) and reported cases of deaths and injuries from self-harm (depicted as a line chart with the left Y-axis). Notably, the dataset reveals a surge in both death and injury cases from September 2019 to October 2019, potentially attributed to changes in the reporting system during the fiscal year transition in Thailand (which starts from October to September).

	date	MS-Pos	MS-Neg	MS-Amb	MS-Neu	ME-Ang	ME-Dis	ME-Fea	ME-Joy	ME-Sad	ME-Sur	ME-Neu	M-NST	M-ST	GH-Death	GH-Injure
0	2017-10-31	0.124349	0.217099	0.002639	0.655914	0.060558	0.001121	0.010423	0.180025	0.042351	0.098291	0.607230	0.734283	0.265717	56.0	224.0
1	2017-11-30	0.122213	0.199027	0.002266	0.676494	0.041806	0.001326	0.016026	0.182476	0.033406	0.104943	0.620016	0.753447	0.246553	45.0	239.0
2	2017-12-31	0.103728	0.244845	0.002444	0.648983	0.057183	0.001756	0.011395	0.179296	0.040314	0.113988	0.596068	0.743003	0.256997	60.0	255.0
3	2018-01-31	0.096537	0.269589	0.002332	0.631543	0.055182	0.001676	0.012206	0.152939	0.024959	0.116959	0.636079	0.746279	0.253721	88.0	336.0
4	2018-02-28	0.093888	0.288119	0.001998	0.615995	0.063627	0.001289	0.011892	0.163508	0.035321	0.118380	0.605983	0.728080	0.271920	60.0	299.0

Fig 2 : Dataset

iv) Pre – processing:

In the data processing phase, the research employs Python libraries such as pandas and numpy to handle and manipulate the datasets effectively. The datasets are initially structured into pandas dataframes, providing a convenient and efficient structure for handling the tabular data. Numpy is then utilized for reshaping operations, allowing for streamlined data manipulation and preprocessing.

Unnecessary columns are dropped from the datasets, ensuring that only relevant information is retained for model training. This step enhances computational efficiency and focuses on features essential for forecasting self-harm trends based on mental signals from social media.

Normalization is applied to the training data, a crucial preprocessing step to standardize numerical features, ensuring they fall within a consistent scale. This normalization helps prevent certain features from dominating others during model training, promoting better convergence and performance.

Following preprocessing, the next step involves extracting training features and labels from the dataset. Features represent the input variables that the model utilizes for prediction, while labels correspond to the target variable, in this case, the number of reported cases of self-harm. This separation ensures that the model learns patterns from the features to make accurate predictions on the target variable, contributing to the overall efficacy of the forecasting framework.

v) Training & Testing:

The dataset is split into training and testing sets to evaluate the model's performance on unseen data, a crucial step in ensuring the generalization of the forecasting model. This division allows the model to learn patterns from the training data and then assess its predictive capabilities on independent, unseen test data.

The training set, typically comprising a larger portion of the dataset, serves as the basis for training the forecasting models. The algorithms learn from the historical patterns and relationships present in this subset of data, enabling them to capture the underlying trends and dynamics related to self-harm cases.

The testing set, on the other hand, remains isolated during the training phase and is reserved for assessing the model's accuracy and performance. The model's ability to make predictions on this unseen data is a

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

critical measure of its generalizability and effectiveness in forecasting self-harm trends. This evaluation ensures that the model does not simply memorize the training data but rather grasps the underlying patterns that can be applied to new, unseen instances.

The split between training and testing sets is often performed randomly to ensure a representative distribution of data, and the results on the testing set provide valuable insights into the model's predictive capabilities and potential real-world applicability.

vi) Algorithms:

ARIMA (AutoRegressive Integrated Moving Average): ARIMA is a time-series forecasting algorithm that models the temporal dependencies in data. It combines autoregression, differencing, and moving averages. In this project, ARIMA is employed to capture the temporal patterns in self-harm trends, leveraging its ability to handle time-dependent variations for accurate forecasting.

Bayesian Ridge: Bayesian Ridge is a probabilistic regression algorithm that incorporates Bayesian principles for regularization. In this project, it is used for forecasting self-harm cases, providing a probabilistic framework to handle uncertainties and improve model robustness against noise in the dataset.

SVR (Support Vector Regression): SVR is a regression algorithm based on support vector machines. In this project, SVR is utilized to model the non-linear relationships between mental signals and self-harm cases. Its flexibility in capturing complex patterns makes it suitable for forecasting in scenarios where the underlying dynamics are intricate.

XGBoost (eXtreme Gradient Boosting): XGBoost is an ensemble learning algorithm that excels in regression tasks. In this project, XGBoost is chosen for its capability to handle complex relationships, feature interactions, and outlier detection. Its boosted tree structure enhances predictive accuracy by combining multiple weak models.

Random Forest: Random Forest is an ensemble learning technique that builds multiple decision trees and merges their predictions. In this project, Random Forest is applied due to its ability to handle non-linear relationships, feature importance analysis, and resistance to overfitting, contributing to robust self-harm predictions.

CatBoost: CatBoost is a gradient boosting algorithm designed for categorical feature support. In this project, CatBoost is selected for its efficiency in handling categorical data from social media, ensuring that information relevant to mental signals is effectively utilized for accurate forecasting of self-harm cases.

Decision Tree: Decision Tree is a simple yet effective algorithm for regression tasks. In this project, Decision Trees are employed to model the decision-making process underlying self-harm trends. Their interpretability and ability to capture non-linear patterns make them valuable components of the forecasting framework.

Voting Regressor: Voting Regressor is an ensemble technique that combines the predictions of multiple regression algorithms. In this project, it is used to aggregate the strengths of various models, enhancing

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

overall prediction accuracy. The diversity in algorithms contributes to a more robust and reliable forecasting system for national-level self-harm trends.

4. EXPERIMENTAL RESULTS

MAE:

Mean Absolute Error (MAE) is a measure of the average size of the mistakes in a collection of predictions, without taking their direction into account. It is measured as the average absolute difference between the predicted values and the actual values and is used to assess the effectiveness of a regression model.

The MAE loss function formula:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

Where:
$\hat{y}_i$ = Predicted value for the ith data point
y_i= Actual value for the ith data point
n = number of observations

RMSE:

Root Mean Squared Error (MSE) is also called Root Mean Squared Deviation (RMSD). RMSE is built on top of MSE, The Root Mean Squared Error (RMSE) is a standard way to measure the error of a model in predicting quantitative data. It is the square root of the Mean Squared Error (MSE) and measures the magnitude of the error in the same unit as the output variable.

RMSE is calculated as the square root of the mean of the squared differences between the predicted values and the actual values using the following formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

Where:
$\hat{y}_i$ = Predicted value for the i^{th} data point
y_i= Actual value for the i^{th} data point
n = number of observations

MAPE:

Mean Absolute Percentage Error (MAPE) is a metric to measure the accuracy of a forecasting model as a percentage. It represents the average absolute percent difference between predicted and actual values for all observations.

MAPE is calculated as the average of the absolute percentage errors between predicted and actual values. The formula for MAPE is as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100$$

Where:
$\hat{y}_i$ = Predicted value for the i^{th} data point
y_i= Actual value for the i^{th} data point
n = number of observations

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

Fig 3: Performance evaluation table for death prediction

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

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

Fig 4: Performance evaluation table for injury prediction

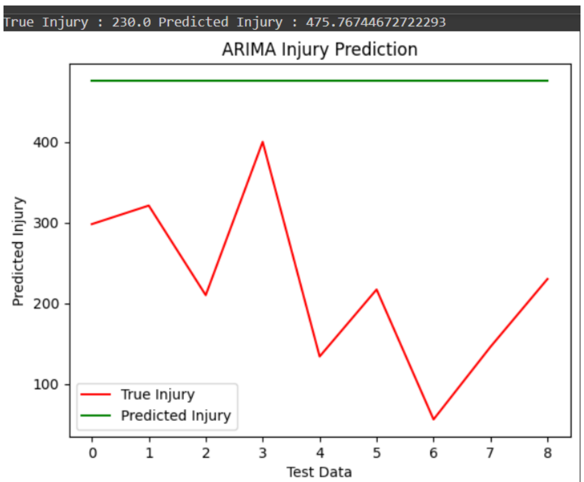

Fig 5 : ARIMA injury prediction graph

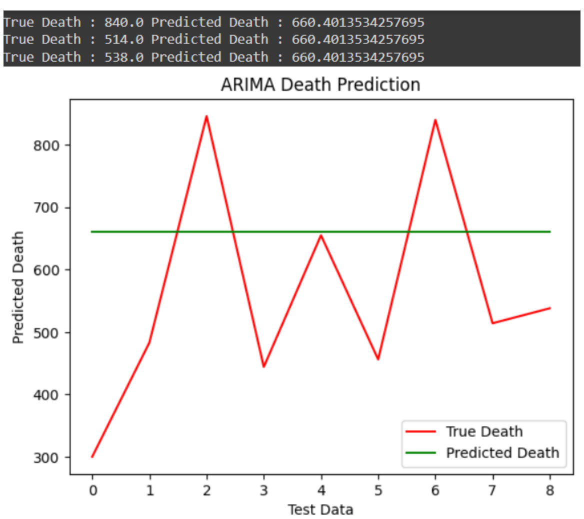

Fig 6 : ARIMA death prediction graph

Fig 7 : Bayesian Ridge injury prediction graph

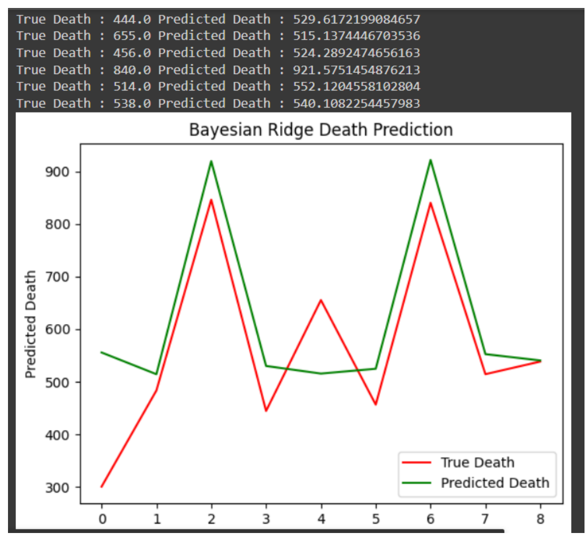

Fig 8 : Bayesian ridge death prediction

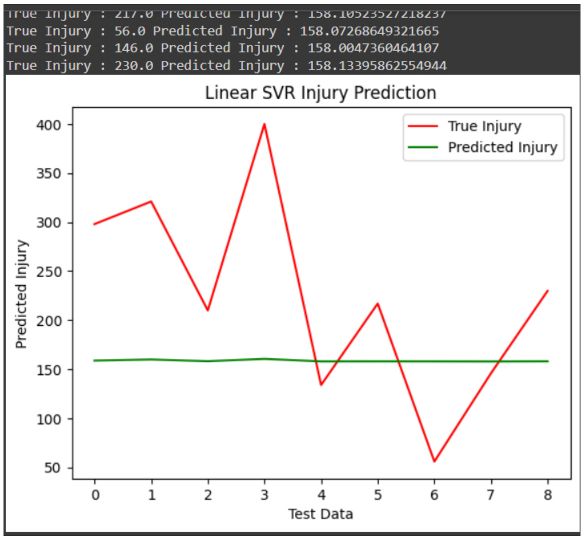

Fig 9 : Linear SVR injury prediction graph

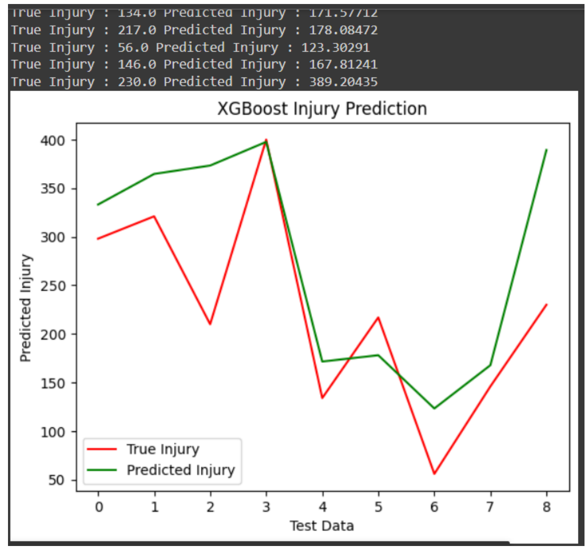

Fig 11: XGBoost injury prediction graph

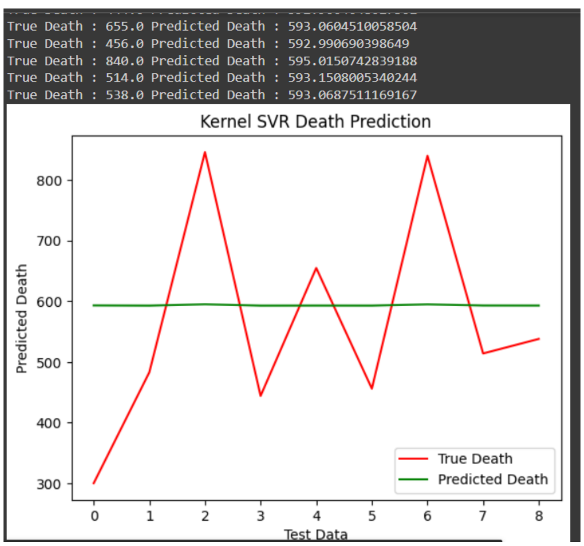

Fig 10 : Linear SVR death prediction graph

Fig 12 : XGBoost death prediction graph

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

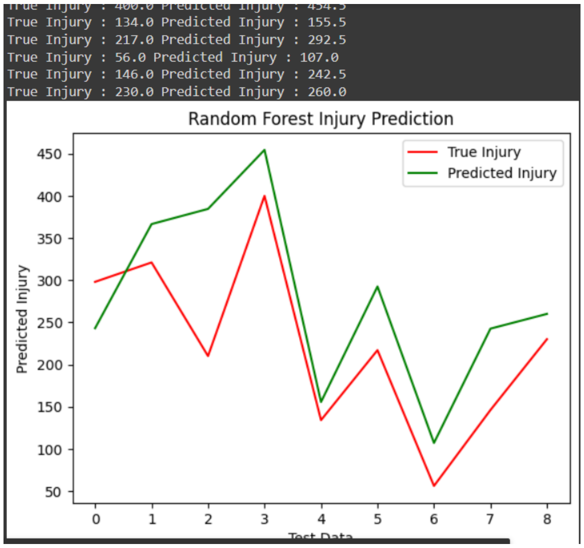

Fig 13 : Random Forest injury prediction graph

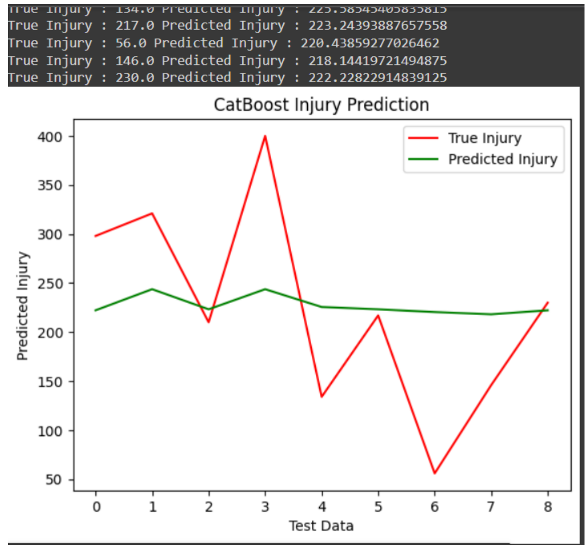

Fig 15 : CatBoost injury prediction graph

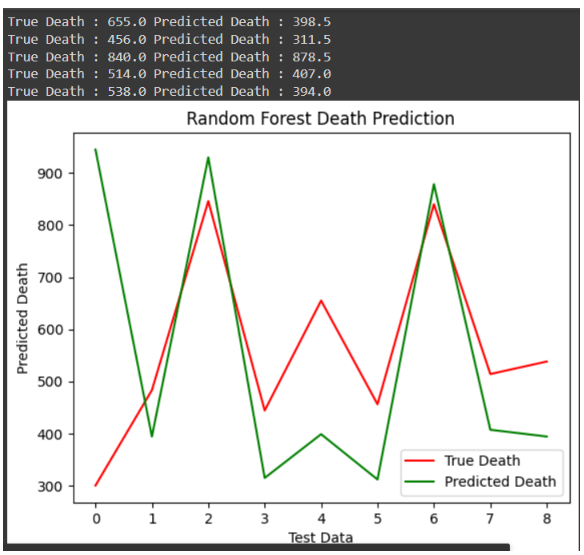

Fig 14 : Random Forest death prediction graph

Fig 16 : CatBoost death prediction graph

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

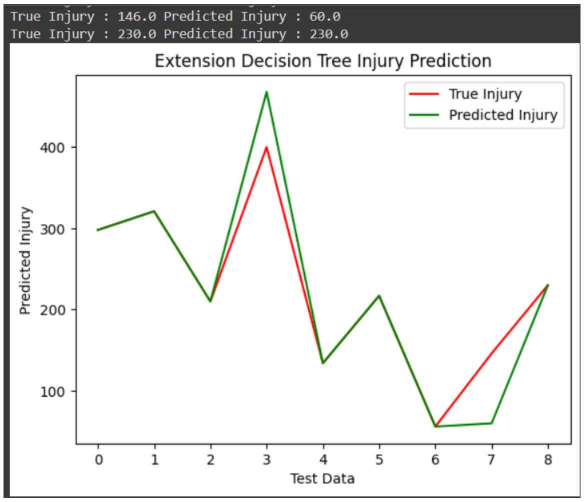

Fig 17 : Extension Decision Tree injury prediction graph

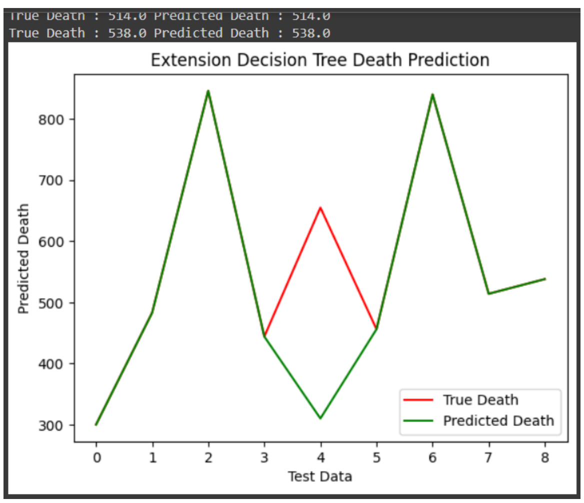

Fig 18 : Extension Decision Tree death prediction graph

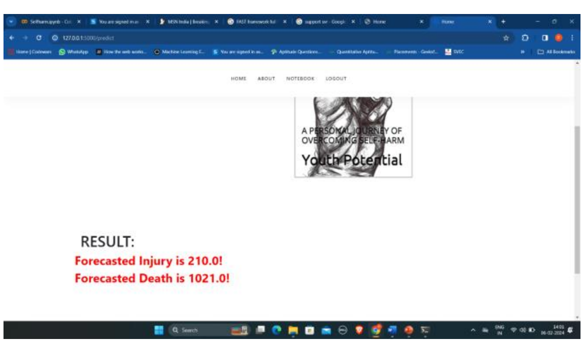

Fig 19 : Predict result for forecasted injury is 210.0 and forecasted death is 1021.0

5. CONCLUSION

In conclusion, this study presents FAST, a pioneering framework for forecasting population-level self-harm trends through the utilization of mental signals extracted from extensive social media data. The research addresses the challenges posed by the unavailability, insufficiency, and delay of ground-truth statistics in certain countries, hindering timely monitoring for proactive policymaking. The framework leverages 12 mental signals extracted from tweets, showcasing its efficacy in improving the forecastability of self-harm death and injury cases in Thailand. Notably, FAST outperforms the traditional ARIMA baseline by 43.56% and 36.48% on average in terms of Mean Absolute Percentage Error (MAPE). The novel approach of utilizing aggregate social media data for nowcasting and forecasting self-harm cases at the national level demonstrates the potential for timely and proactive intervention by policymakers and public health stakeholders. While the experimental results are promising, the study acknowledges the need for further enhancements. We further research avenues may explore additional techniques, such as Decision Tree and Voting Regressor, to continuously refine the framework and contribute to the ongoing efforts in addressing the escalating trends of self-harm associated with technological advancements and rapid urbanization in developing countries.

6. FUTURE SCOPE

Future research endeavors could expand the scope by exploring diverse online media sources, including news articles, various social media platforms, and

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

multimedia content like videos or photos, to enrich the dataset and enhance forecasting accuracy. Additionally, the integration of deep learning techniques could unlock more intricate patterns within the data. Investigating self-harm incidents at a granular level, such as regional or demographic-specific scales, holds promise for tailoring management approaches, enabling the development of targeted interventions tailored to specific localities and demographics, thereby advancing precision in preventive strategies.

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] E. Iob, A. Steptoe, and D. Fancourt, ''Abuse, self-harm and suicidal ideation in the U.K. During the COVID-19 pandemic,'' Brit. J. Psychiatry, vol. 217, no. 4, pp. 543–546, Oct. 2020.

[37] M. I. Islam, R. Khanam, and E. Kabir, ''Depression and anxiety have a larger impact on bullied girls than on boys to experience self-harm and suicidality: A mediation analysis,'' J. Affect. Disorders, vol. 297, pp. 250–258, Jan. 2022.

[38] S.-P. Jun, H. S. Yoo, and S. Choi, ''Ten years of research change using Google trends: From the perspective of big data utilizations and applications,'' Technol. Forecasting Social Change, vol. 130, pp. 69–87, May 2018.

[39] S. Kandula, M. Olfson, M. S. Gould, K. M. Keyes, and J. Shaman, ''Hindcasts and forecasts of suicide mortality in U.S.: A modeling study,'' PLOS Comput. Biol., vol. 19, no. 3, Mar. 2023, Art. no. e1010945.

[40] R. C. Kessler, R. M. Bossarte, A. Luedtke, A. M. Zaslavsky, and J. R. Zubizarreta, ''Suicide prediction models: A critical review of recent research with recommendations for the way forward,'' Mol. Psychiatry, vol. 25, no. 1, pp. 168–179, Jan. 2020.