A Framework for Predicting Self-Harm Using Social Media Signals

[1] Kanugula Saikumar, [2] Dr. A. Arjuna Rao, [3] Kotyada Appala Ramu
Student, Professor, Assistant Professor
Dept. Of Master of Computer Applications,
Miracle Educational Society Group of Institutions

ABSTRACT

Mental health issues such as self-harm can be highly dangerous on a personal and societal level. Economic tolls are also considerable. People self harm due to psychological discomfort or deep distress. Automated processes that solely look at historical information have shown, time and time again, to completely miss consideration of emotions. Emotions play an essential role in the manner in which individuals function. Untapped socioeconomic potential of nations can be provided through interventions targeted at social media analysis, including social indicators in data analysis, and refining self-harm prediction models on a national scale. This model has successfully conducted emotion recognition, transforming emotions such as sadness, anxiety, and experiences of suicidal thoughts into time series data from various social media platforms enabling the prediction and detection of self-harm patterns with adequate time. Subsequent testing showed significant improvements to the framework when implemented with Decision Trees and XGBoost in conjunction with machine learning algorithms. Thailand proved to be an excellent candidate, showcasing over 40% improvement relative to previous frameworks utilizing conventional methodologies.

Keywords: self-harm, Decision Trees

INTRODUCTION:

The lack of policies and technological advancements is where this phenomenon stems from. In developing countries where self-harming individuals account for over 77% of an entire nations suicide rate self-harm emerges as a prevailing phenomenon needing further study alongside mental health concerns. Today's frameworks depend on hospital records and government documents which are expensive and slow to gather. With social media serving as a digital outlet for self-expression, there is an opportunity to monitor trends for real-time mental health assessment. Studies indicate that people post about feeling desperate, anxious, and sad on Facebook and Twitter. With social media text analysis, population mental state estimation can be achieved. The FAST framework seeks to capture this data by translating emotional signals into predictive algorithms for self-harm injuries and fatalities, enabling faster response measures and policy implementation to mitigate emerging mental healthcare challenges.

GAP IDENTIFIED BASED ON LITERATURE SURVEY:

There is a noted lack of research regarding the prediction and analysis of self-harm through dynamic and real-time resources at the national level. Existing predictive models like ARIMA and Google Trends have proven to be ineffective due to reliance on outdated historical data or opaque algorithms that are not behaviorally driven.

Literature suggests that while suicide-related search volumes might provide some correlation with public sentiment, they fail to offer consistent accuracy or represent actual self-harming populations. Furthermore, previous studies often neglect the vast and openly available behavioral indicators embedded in social media content.

Another significant gap lies in the granularity and timeliness of official data. National databases are frequently outdated and lack the nuance needed for effective early intervention strategies. Attempts to forecast using clinical or economic indicators do not account for the everyday expressions of distress that individuals post online. While some research has focused on individual detection of suicidal ideation through social media, the application of these signals to forecast trends at the national level remains largely unexplored. The FAST framework addresses this gap by aggregating emotional data from social networks, transforming them into scalable time-series predictors for nationwide self-harm trends.

PROBLEM STATEMENT:

Forecasting national-level self-harm trends remains a critical yet complex challenge. Traditional data sources are delayed, sparse, and often unreliable for proactive intervention. The problem is further compounded by the absence of scalable systems to interpret population-wide mental health signals in real time.

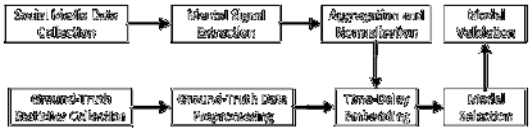
Key Challenges:

- Reliance on historical or incomplete datasets
- Lack of real-time behavioral data for forecasting
- Difficulty in extracting emotional indicators from social media text
- Limited model generalizability across regions and languages
- High variance in individual expression and mental health disclosure

PROPOSED METHOD:

The proposed approach revolves around a framework called FAST, which utilizes social media data to predict self-harm trends. The system uses language-agnostic models to fetch mental health related posts containing emotional pointers like ‘sad’, ‘anxious’, or ‘suicidal’. These posts are processed for cognitive indicators of sadness, anxiety, and even suicidal thoughts. Emotional data is collected and synchronized with recorded instances of self-harm to create multivariate time-series datasets. Machine learning models which include ARIMA, SVR, XGBoost, CatBoost, and Decision Trees are trained and tested to assess which of the best predictors out of them is most effective. The most accurate model is chosen for projection at the national level. This model attempts to minimize the scope of extrapolated self-harm behaviors by offering a lucid and rapidly adaptable way to forecast potential suicide trends.

ARCHITECTURE:



DATASET:

The dataset used combines two primary sources: (1) official health department records of self-harm injuries and deaths in Thailand, and (2) emotional signals derived from social media posts. Social media data was gathered using predefined mental health-related keywords and processed through language models to extract sentiments and suicidal tendencies. The final dataset consists of multivariate time-series entries, where each row represents a time snapshot of aggregated emotions and corresponding self-harm incidents. Missing values were handled using zero-fill methods, and the dataset was shuffled and split into training and testing sets to build robust machine learning models.

METHODOLOGY:

Step 1: Data Collection

Extracted large-scale, public social media posts using mental health-related keywords.

Gathered historical self-harm statistics (injuries and deaths) from Thailand’s health department.

Step 2: Preprocessing

Cleaned and normalized social media text data.

Extracted emotional features (sadness, anger, suicidal thoughts) using NLP models.

Handled missing values using zero-filling and standardized datasets.

Step 3: Feature Engineering

Converted individual emotional posts into aggregated mental signals on a daily/weekly basis.

Created multivariate time-series datasets combining mental signals and historical self-harm incidents.

Step 4: Dataset Splitting

Divided the dataset into 80% training and 20% testing sets using train_test_split.

Separate models were created for predicting injuries and deaths.

Step 5: Model Selection

Implemented several ML regressors: ARIMA, Bayesian Ridge, SVR, XGBoost, Random Forest, CatBoost, and Decision Tree.

Chose ARIMA for time-series only models, and others for multivariate regression.

Step 6: Training and Evaluation

Trained models using training datasets.

Evaluated models using MAE, RMSE, and MAPE metrics.

Plotted predicted vs. true values to visualize forecasting accuracy.

Step 7: Model Comparison

Compared model performance across metrics for both injury and death forecasts.

Identified Decision Tree and XGBoost as top performers based on lower RMSE and MAE.

Step 8: Final Prediction

Applied best-performing model (Decision Tree) on test dataset to forecast unseen data.

Forecasted both injury and death counts based on real-time social signals.

Step 9: Visualization

Generated comparative charts of model performance.

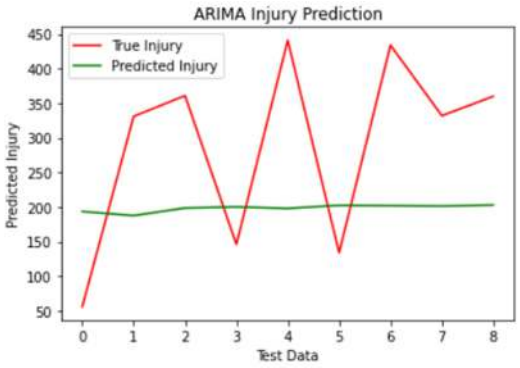
Displayed test data forecasts for policymakers and health stakeholders.

Step 10: Interpretation and Application

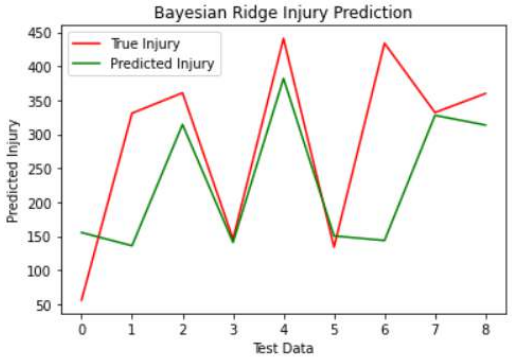
Used results to identify peak periods and regions of rising self-harm risk.

Framework aids in targeted intervention planning, including mental health support rollouts.

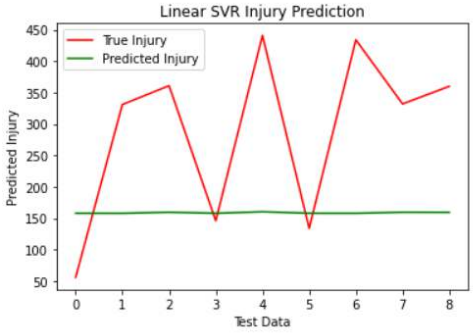
RESULTS:



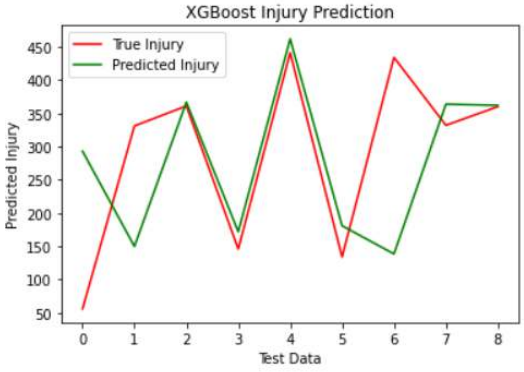
ARIMA Injury MAPE : 25342.089430205095



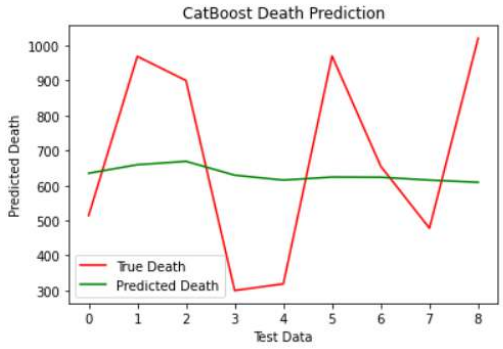
Bayesian Ridge Injury MAPE : 15581.43050507478



Linear SVR Injury MAPE : 34053.76420271253



XGBoost Death MAPE : 29639.06855095095



Extension Decision Tree Injury MAPE : 9303.111111111111

Forecasted Injury form test data

Test Data = ['2019-04-30' 0.1001397518460463 0.2199769598307869 0.005684500764858001

0.6741987875583086 0.0387622518932598 0.0016052576910728

0.0078279919170553 0.143812203735529 0.0192914203697758

0.0881664180091027 0.7005344563842043 0.7598156786462956

0.2401843213537044]=====> Forecasted Injury : 157.0 Forecasted Death : 550.0

Test Data = ['2019-05-31' 0.1055472075148101 0.2527987790007002 0.0055070844708085

0.6361469290136812 0.04802964384898 0.001667859868302 0.0095823269792068

0.1715928848468637 0.0302102919541495 0.099568087232218

0.6393489052702799 0.7391766121989788 0.2608233878010212]=====> Forecasted Injury : 134.0 Forecasted Death : 538.0

Test Data = ['2019-06-30' 0.1045612288819541 0.2629936511352631 0.0056628645216829

0.6267822554610998 0.0522099967717636 0.0013181964919832

0.0121462391047024 0.1625955019907457 0.0389809534057893

0.09796352092973198 0.6347855913052836 0.7278462283439148

0.2721537716560852]=====> Forecasted Injury : 154.0 Forecasted Death : 514.0

CONCLUSION

This project introduces a novel approach to forecasting national self-harm trends using real-time social media data. By leveraging emotional indicators and training advanced machine learning models, the FAST framework demonstrates significant improvement over traditional forecasting methods. The use of Decision Trees and XGBoost allowed for more accurate predictions, enabling earlier identification of rising mental health risks. The integration of social signals into public health forecasting not only improves prediction accuracy but also offers a scalable and timely solution for national-level mental health monitoring. With proper implementation, such systems can support timely interventions and reduce self-harm occurrences on a broader scale.

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, ‘‘TheGatedTabTransformer. 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.