International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Social Media Insight: Predicting Self-Harm Trends

PRATHYUSHA BURRA * and GNR PRASAD

Department of MCA, Chaitanya Bharathi Institute of Technology, India.

International Journal of Science and Research Archive, 2024, 13(01), 3068–3080

Publication history: Received on 11 September 2024; revised on 20 October 2024; accepted on 23 October 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.13.1.2006

Abstract

Poisoning or injuring oneself can cause death or injury without intent. This is self-harm. Self-harm harms the perpetrators and the nation's economy. Studies have linked increased urbanization in emerging nations to rising self-harm and new technology. National self-harm patterns may be important for policymakers and public health experts to predict. This would allow them to immediately fix issues or avoid disasters. Studies have forecasted population-level self-harm patterns in many nations using simple statistics. In certain nations, prior data may be scarce or insufficient to make reliable projections. This makes it tougher to FAST assess and anticipate national self-harm. This essay proposes FAST, a technique that uses mental indicators from social media to forecast national self-harm trends. These markers can represent community mental health and help forecast self-harm patterns. Language-agnostic algorithms are taught to first identify mental indicators in gathered social media communications. These signals form multivariate time series. The time-delay embedding method then places these events in time. Last, many machine learning regressors are tried for future prediction. Tweets with 12 mental indicators may predict self-harm-related fatalities and injuries, according to a Thailand research. The suggested method predicted self-harm deaths and injuries. Better than ARIMA baseline by 43.56% and 36.48%. We think our research is the first to predict national self-harm trends using social media data. The findings improve self-harm prediction and establish the framework for new social network-based apps that forecast socioeconomic aspects. We used the top machine learning algorithms, Decision Tree and Voting Regressor. We found less MAE mistakes with these techniques.

Keywords: Self-harm; Nowcasting; Forecasting; Online social networks; Cross-lingual text classification

1. Introduction

Self-harm happens when somebody purposefully toxins or damages himself, no matter what the rationale or level of distress. It can harm or kill them. Many individuals, especially in less fortunate countries, harm or commit suicide. As per a new report, most of suicides (around 77%) happened in countries with low, center, or top level salaries. Individuals accept this propensity is connected with how FAST individuals around there are moving to urban communities and taking on new innovation. Self-harm is developing progressively far reaching, which causes agony and misfortune for individuals as well as has long haul adverse results for the economy, for the most part by bringing down long haul work creation. Having the option to distinguish and expect designs in self-harm all through a local area could help public legislators and general wellbeing accomplices in FAST surveying what is going on and executing procedures to forestall or lessen the quantity of projected passings. In the wake of understanding that drastic actions planned at forestalling significant plagues have caused psychological well-being worries in specific individuals and are supposed to fuel self-harm, specialists might rethink the approaches that are creating these issues. Particular general wellbeing endeavors, like versatile mental units or hotlines, may help burdened populaces. The best way to follow public self-harm patterns is through medical clinic and medical services organization managerial records. This approach requires a critical venture of cash, work force, and time, consequently information is just provided irregularly and late on

* Corresponding author: Burra Prathyusha

occasion. Insights that are too wide or come past the point of no return might be inadequate for settling on informed strategy choices.

This article presents FAST, a technique for anticipating public self-harm designs utilizing cerebrum signs removed from web-based entertainment information. Language-rationalist models dissect messages and convert them into multivariate time series changed by time-delay installing. Decision Tree and Casting a ballot are two ML regressors that beat customary methodologies in determining passing and self-harm. This furnishes legislators with indispensable data.

The expansion in self-harm episodes is connected to further developed innovation and more people moving to urban communities in arising countries. This makes it harder to figure and appreciate what will occur at the public level. Standard methodologies that depend on authentic realities may not be adequate. This study proposes FAST, which use mental pieces of information gained from web-based entertainment to gauge self-harm propensities. This is an endeavor to address the interest for further developed general wellbeing expectation models.

2. Literature survey

[37] This study inspects the perplexing connection between tormenting, self-damage, and self-destruction. It analyzes how despairing and nervousness impact these connections diversely for people. Aristocrat and Kenny's procedure is utilized to analyze the effects on guys and young ladies in 2522 Australian teens matured 12-17. 53.2% of harassed were females, 46.8% young men. This suggests young ladies are more inclined to be miserable, stressed, self-harm, or self-destructive. Logistic regressions and the Sobel test show that distress is a main consideration in the connection between tormenting, self-mischief, and self-destruction in people. In any case, stress issues solely impact young ladies. These discoveries uncover that tormenting influences females, featuring the requirement for orientation explicit security mediations. Cross-sectional strategies and self-detailed information might produce issues, so don't derive circumstances and logical results ends. The outcomes show the need to FAST foster explicit projects to address the mind boggling ways harassing influences young men and young ladies' psychological well-being, with an accentuation on how despairing and uneasiness are involved.

[21] This drawn out accomplice study assesses 15,644 general non-mental hospitalizations of discouraged, bipolar, and crazy patients. To survey self-destructive and self-harm readmission risk after broad hospitalization. The Classification and Regression Tree strategy anticipated self-destruction endeavor and self-harm in the following year utilizing organized electronic wellbeing record information from a southwestern US metropolitan wellbeing framework from 2006 to 2017. Information from the earlier year and file hospitalization were used. An AUC of 0.86 showed the program's great grouping assessment and forecast capacity. particular gatherings showed unmistakable self-destruction related conduct. The best rates were seen among hospitalized self-destruction attempters. Most dangers can be made sense of by a mix of indicators, including liquor use and a moderate clinical bleakness, and being more youthful than 55 with a low dreariness. These discoveries exhibit the capability of straightforward AI calculations to help clinicians in direction, asset portion, and counteraction for people with serious mental illnesses in everyday emergency clinics.

[32] The CLPsych 2021 Shared Undertaking shows a tweet-based suicide risk detection system. The procedure utilizes Self-Harm Topic Model (SHTM), in view of emotional wellness studies connecting self-destructive ideation to self-harm. This model utilizes a self-harm jargon and Dormant Dirichlet Distribution to describe tweets with respect to self-harm across time. SHTM permits further examination of everyday tweets to represent how feelings and subjects advance. From that point onward, a profound gaining network gains self-destruction forecast from these differences. The framework utilizes point displaying and DL out how to get a handle on clients' self-harm thoughts and feelings. This will assist with determining self-destruction risk all the more precisely. Virtual entertainment information and refined displaying can anticipate self-destruction risk, as indicated by the review. It likewise represents how computational methodologies further develop psychological well-being observing and avoidance.

[22] This orderly survey and meta-investigation looks at what makes convicts bound to harm themselves, a significant medical problem. The assessment of 35 examinations from 20 countries and 663,735 convicts uncovered that self-destructive factors drove detained self-harm the most. Psychological wellness hardships are vital since present or ongoing self-destructive considerations, a past filled with self-destruction ideation, and self-harm are unequivocally connected. A psychological maladjustment, serious despondency, or marginal behavioral condition builds the gamble of self-harm. Jail explicit outer factors including being separated from everyone else, ignoring the principles, and being physically or genuinely abused were likewise fundamental. In any case, sociodemographic and criminological qualities just fundamentally impacted jail self-harm. A thorough, jail wide system to self-harm counteraction is required because of the enormous scope of chance variables. The review exhorts that psychological well-being, social consideration, and

law enforcement organizations team up on populace wide and designated intercessions. These actions ought to consider the intricate linkages among clinical and care related regions to lessen prison self-harm.

[36] This study analyzed UK misuse, self-harm, and self-destructive or self-harming ideation utilizing the Coronavirus Social Review, which remembered 44,775 people for the principal month of the pestilence. These unsavory experiences were more normal for ladies, BAME people, and those with cash, joblessness, hindrances, long haul illnesses, mental problems, or Coronavirus analyze. Mental drugs were the most famous kind of help, despite the fact that less than half of those in need utilized official or casual organizations. These patterns exhibit what the pandemic has meant for various socioeconomics and monetary levels. Individuals are more disposed to abuse, harm themselves, and think about self-destruction during cultural distress. To tackle this issue, we want designated intercessions and backing systems, particularly for distraught populaces, as per this review. The way that current treatment isn't completely used highlights that it is so essential to make emotional well-being assets easier to find and more available to diminish the pandemic's adverse effects on psychological wellness.

3. Methodology

3.1. Proposed Work

This paper proposes Quick, a method that utilizes mental markers from social media to forecast national self-harm trends. These markers can address local area emotional well-being and assist with forecast self-harm designs. Language-rationalist calculations are educated to initially recognize mental pointers in assembled social media communications. These signs are joined into multivariate time series. Then, at that point, the time-delay implanting approach inserts these events in time. At last, a few machine learning regressors are tried for future prediction. We attempted the Decision Tree and Voting Regressor, two top ML strategies. It makes less MAE error than others.

3.2. System Architecture

A nationwide self-harm prediction approach utilizing social networks has a few key parts. In the first place, distinguish significant characteristics in the "self-harm and mental-signals" assortment. From that point onward, data processing methodology clean and set up the information for investigation.

Splitting the data into training and testing sets works on model testing. Anticipating models incorporate ARIMA, Bayesian Edge, SVR, XGBoost, Random Forest, CatBoost, Decision Tree, and Voting Regressor ensemble.

Each model is prepared and fabricated utilizing various philosophies and hyperparameter tweaking. Each model is assessed utilizing accuracy, precision, memory, and F1 score to predict self-harm trends.

The framework design accentuates adaptability and utilizes numerous strategies to increment forecast accuracy. The Voting Regressor consolidates the best of different sorts. The strategy empowers for a countrywide examination of self-harm designs utilizing social media data to further develop estimates.

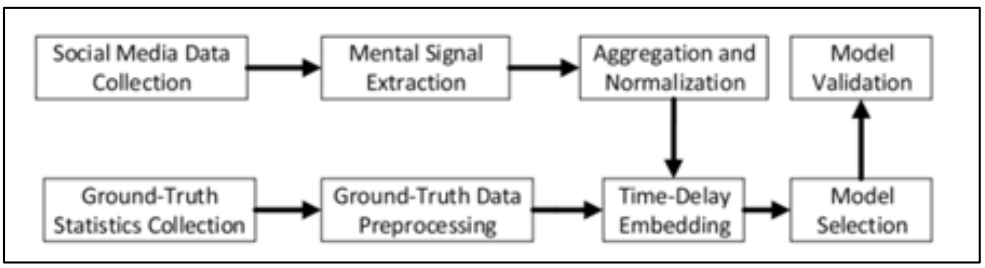

Figure 1 System Architecture

3.3. Dataset Collection:

The data collection used in this study is significant to understanding how social media mental signs anticipate public self-harm rates. The contextual analysis blends social media data in with genuine data in regards to self-harmers in Thailand.

In this examination, two informational indexes are utilized. An irregular example of 4.9 million tweets from October 2017 to January 2021 was gathered utilizing the Twitter API. Just the timestamp and text of each tweet are put away to guarantee consistency across informal communication stages. Since it needn't bother with language to draw mental pointers, the estimating approach might be applied with multivariate time series information. This improves on use in numerous dialects and areas.

The second arrangement of data contains Thai self-harm passings and wounds from earlier months. This information came from Thailand's Ministry of Public Health's Mental Health Department. It conveys solid model approval information. The dataset shows month to month tweets (bar graph on right Y-pivot) and self-harm fatalities and wounds (line diagram on left Y-hub). Passings and wounds rose from September to October 2019. Thailand's monetary year progress from October to September might have impacted case announcing.

	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

Figure 2 Dataset

3.4. Pre – processing:

Data processing includes utilizing Python devices like pandas and NumPy to control datasets. Beginning datasets are assembled into pandas data frames for simple and quick plain information control. For twisting position, NumPy improves on data preparation and processing.

Dropping unnecessary factors from datasets saves simply fundamental information for model training. This stage speeds up the PC and focuses on characteristics that expect self-harm designs in light of social media mental signs.

Normalizing training data guarantees that all numbers are a similar size all through arranging. Normalization keeps a few qualities from turning out to be more critical during model training, further developing intermingling and execution.

Remove training features and labels from the dataset subsequent to cleaning. Names address the point variable, which is self-hurt occasions, while highlights uncover the model's feedback components. This gap assists the model with gaining designs from highlights to appraise the objective variable precisely. This improves foreseeing generally speaking.

3.5. Training & Testing:

The data is isolated into training and testing sets so the model might be tried on new data. This is essential to guaranteeing the gauging model works in all circumstances. This part permits the model gain patterns from training data and estimate on new test information.

A bigger part of the information is used to prepare estimating models. The calculations gain from this information's authentic examples and linkages. This assists individuals with recognizing self-harm patterns and changes.

In any case, the testing set is kept separate all through preparing and exclusively used to evaluate model execution. What the model can foresee from this new information is a basic sign of its ability to figure self-hurt designs. This test guarantees that the model training data and grasps designs for new models.

To disseminate information evenhandedly, preparing and it are by and large arbitrarily separated to test sets. The testing set shows how well the model predicts and how it could work, in actuality.

3.6. Algorithms

- **ARIMA (AutoRegressive Integrated Moving Average):** The time-series anticipating technique known as ARIMA (AutoRegressive Integrated Moving Average) models the worldly connections between data of interest. It utilizes moving midpoints, differencing, and autoregression. To make precise conjectures, this study utilized ARIMA to follow self-harming trends across time.
- **Bayesian Ridge:** Bayesian regularization is utilized in this probabilistic regression method. It is utilized in this work to gauge examples of self-harm, giving a probabilistic system to deal with vulnerability and work on model power against commotion in the dataset.
- **Support Vector Regression:** Support vector machines are utilized in SVR to do regression. The non-linear correlations between self-harm and mental signs in this study are demonstrated utilizing SVR. As a result of its capacity in catching multifaceted examples, it can expect complex elements.
- **XGBoost (eXtreme Gradient Boosting):** Regression is an area of strength for a for the ensemble learning technique known as XGBoost (eXtreme Gradient Boosting). XGBoost is utilized in this venture to distinguish anomalies and handle complex element communications and connections. Prediction accuracy is expanded by consolidating numerous feeble models in its boosted tree structure.
- **Random Forest:** The forecasts of many decision trees are consolidated in this ensemble learning method. This study utilizes Random Forest to foresee self-harm as a result of its capacity to deal with non-linear relationships, feature importance investigation, and protection from overfitting.
- **CatBoost:** CatBoost utilizes gradient boosting to support categorical features. This study utilizes CatBoost in light of the fact that it actually oversees class social media data, guaranteeing that psychological sign data is used to estimate examples of self-harm unequivocally.
- **Decision Tree:** This straightforward yet powerful relapse strategy is called Decision Tree. This study utilizes Decision Trees to address dynamic connected with self-harm. They're ideally suited for anticipating in view of their interpretability and capacity to get non-straight examples.
- **Voting Regressor:** The predictions of numerous regression calculations are joined in this ensemble method. To increment forecast accuracy, this undertaking joins the qualities of the models. Predictions of national trends in self-harm are improved by diversifying algorithms.

4. Experimental results

4.1. MAE

Mean Absolute Error (MAE) is a proportion of the typical greatness of mistakes in a bunch of predictions, without respect for heading. It is determined as the normal outright contrast among anticipated and genuine qualities and used to assess the viability 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 i^{th} data point
y_i = Actual value for the i^{th} data point
n = number of observations

4.2. RMSE

MSE is otherwise called RMSD. The Root Mean Squared Error (RMSE) is a commonplace strategy for estimating model blunder in gauging quantitative information. The square base of the Mean Squared Error (MSE) measures mistake extent in a similar unit as the result variable.

This equation computes RMSE as the square root of the mean of the squared inconsistencies among expected and genuine qualities:

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

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

4.3. MAPE

The Mean Absolute Percentage Error (MAPE) is a measurement that actions an estimating model's exactness as a rate. It addresses the typical outright rate distinction among expected and real qualities across all perceptions.

MAPE is inferred by averaging the outright rate mistakes among anticipated and genuine qualities. The recipe for MAPE is as per the following:

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

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

Figure 3 Performance evaluation table for death prediction

	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

Figure 4 Performance evaluation table for injury prediction

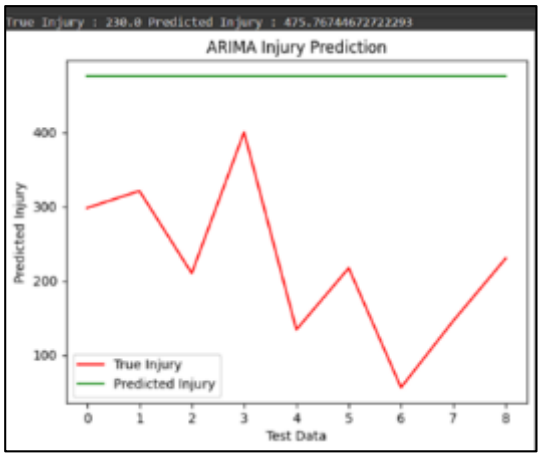

Figure 5 ARIMA injury prediction graph

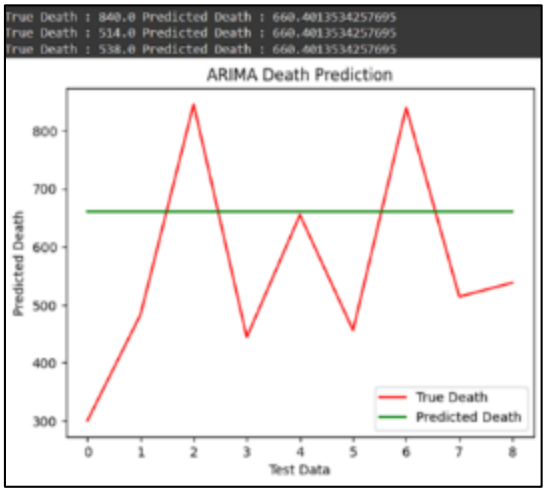

Figure 6 ARIMA death prediction graph

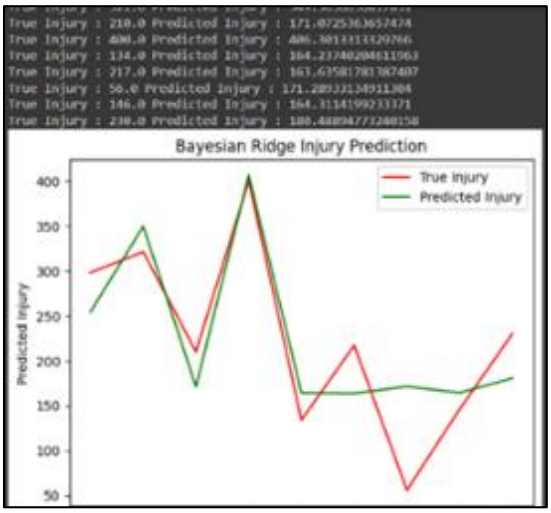

Figure 7 Bayesian Ridge injury prediction graph

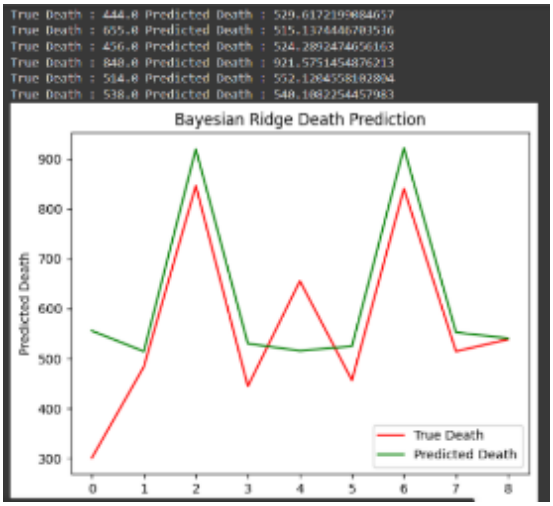

Figure 8 Bayesian ridge death prediction

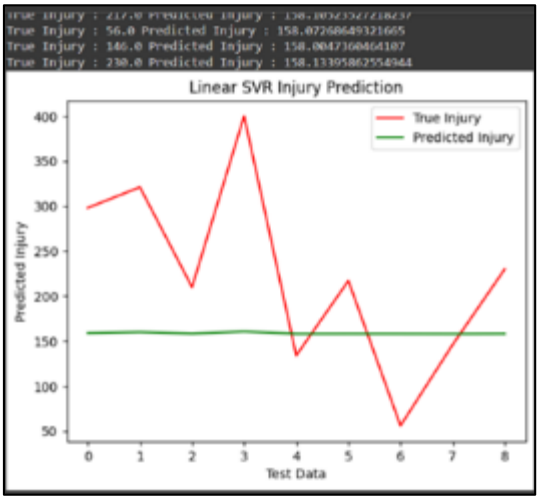

Figure 9 Linear SVR injury prediction graph

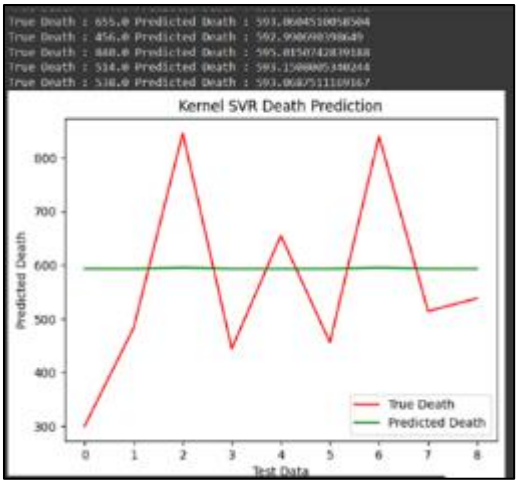

Figure 10 Linear SVR death prediction graph

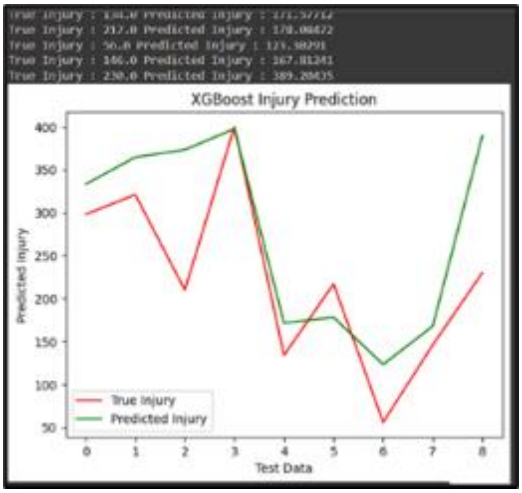

Figure 11 XGBoost injury prediction graph

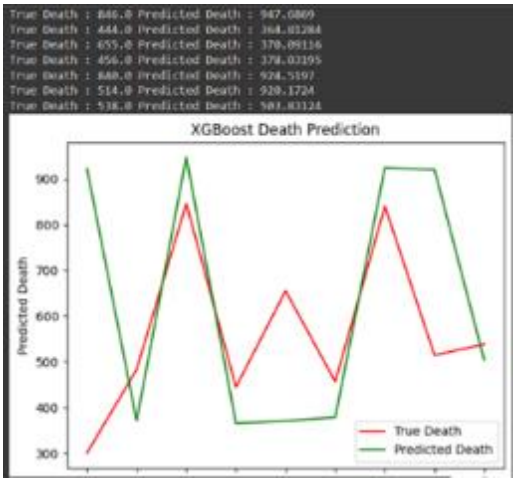

Figure 12 XGBoost death prediction graph

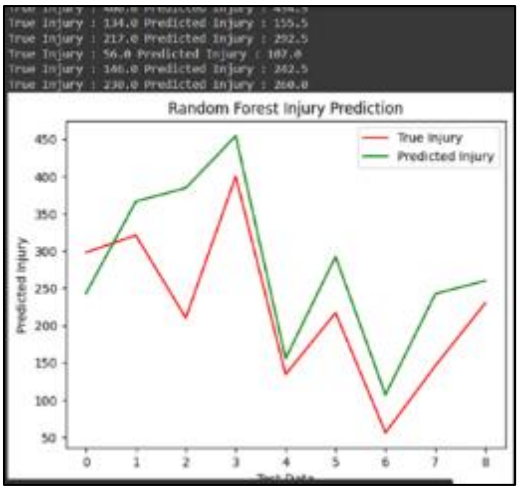

Figure 13 Random Forest injury prediction graph

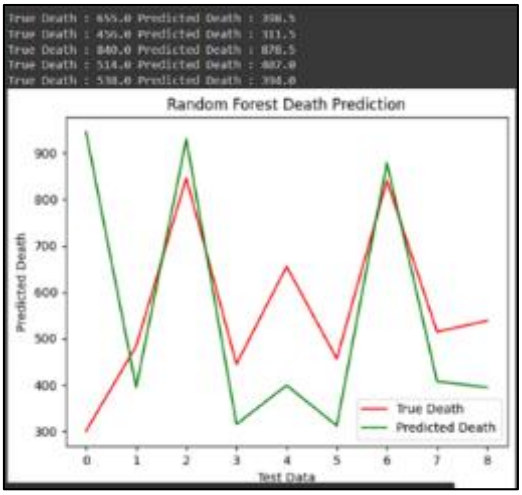

Figure 14 Random Forest death prediction graph

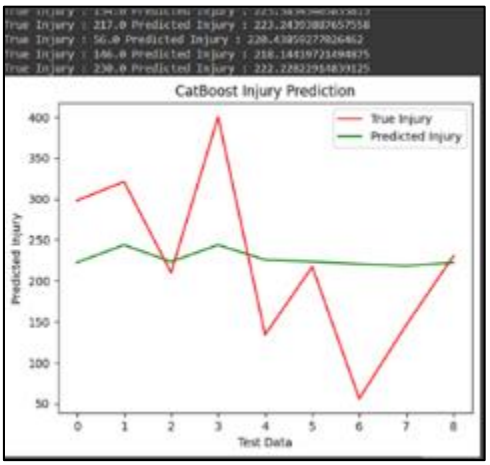

Figure 15 CatBoost injury prediction graph

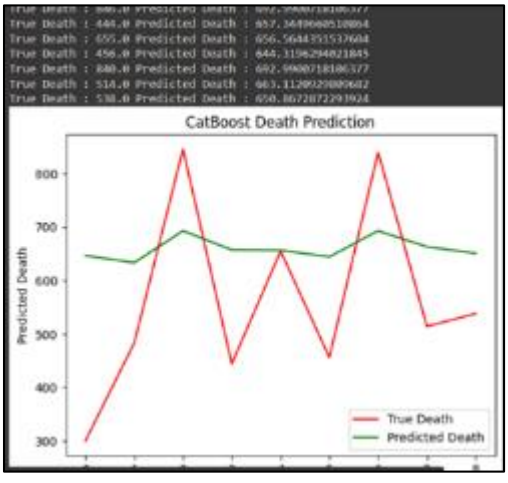

Figure 16 CatBoost death prediction graph

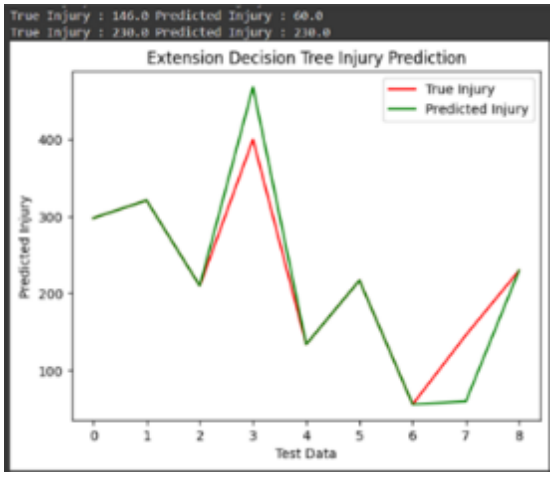

Figure 17 Extension Decision Tree injury prediction graph

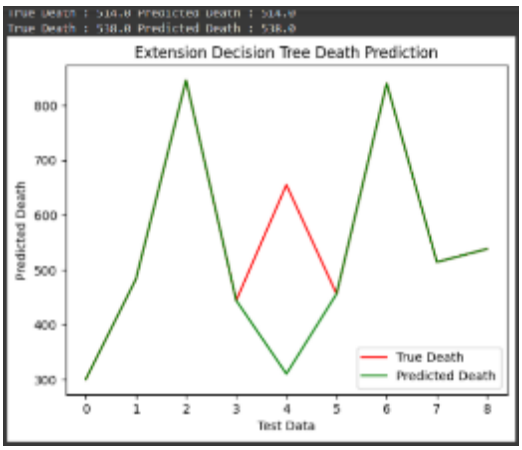

Figure 18 Extension Decision Tree death prediction graph

Figure 19 Predict result for forecasted injury is 210.0 and forecasted death is 1021.0

5. Conclusion

FAST, an original method to forecast community-wide self-harm designs utilizing mental pieces of information from social media data, finishes up this examination. The paper analyzes issues when ground-truth information are not many, deficient, or postponed in certain nations. This makes it harder to screen things rapidly enough to make powerful

strategy decisions. The technique utilizes twelve tweet-based mental signs to predict Thai self-harm passing and injury events. FAST beats ARIMA by 43.56% and 36.48% in Mean Absolute Percentage Error. The better approach for utilizing enormous social media data to expect and nowcast cross country self-harm cases recommends government officials and general wellbeing gatherings can act rapidly and effectively. In spite of positive testing results, the examination surrenders extra work is expected to further develop them. We are investigating ways to deal with utilize Decision Tree and Voting Regressor to work on the structure and lessen the rising propensity of self-harm brought about by new innovation and fast urbanization in arising countries.

Future scope

Future examination could incorporate looking at other web-based media designs, including news stories, social media stages, and sight and sound substance like pictures and recordings. Thusly, the assortment would develop and the conjectures would get more precise. Deep learning methods can possibly reveal significantly more perplexing examples in the information. Modifying the executive's arrangements might profit from a nearer assessment of self-harm episodes at a nittier gritty level, like a geological or segment level. This will improve the precision of preventive projects by empowering the advancement of fitted medicines that are specific to specific districts and socioeconomics.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

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. Chong suvivatwong, 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 Gated Tab Transformer. 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.37 10.18653/01/2020.acl-main.372. doi: 10.18653/01/2020.acl-main.372.

[21] J. B. Edgcomb, T. Shaddox, G. Hellemann, and J. O. Brooks, "Predicting suicidal behavior and self- Charm 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.