IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Social Sentinel: Predicting National Self-Harm Trends Trough Social Networks

D.Gayathri[1],Mr.G.Lokesh[2]

[1]PG student, vemu institute technology,P.Kothakota,

[2]Assistant Professor, vemu institute technology,P.Kothakota

ABSTRACT

Since this study delves into the profound impacts of self-harm on individuals and economies, emphasizing the inadequacy of traditional statistics in tracking national trends. Introducing the innovative FAST framework, it harnesses social media data to forecast self-harm incidents. By training language models to discern mental health signals from online messages, this method transforms them into insightful time series data. Using machine learning regressors, the framework demonstrated superior forecasting accuracy. In a Thai case study, it surpassed conventional methods by over 40%. Additionally, incorporating the Decision Tree algorithm enhanced accuracy, reducing Mean Absolute Error compared to other algorithms. This research pioneers a transformative approach to predict nationwide self-harm trends and potentially forecast socioeconomic factors using social media analytics.

Keywords: Self-Harm, Social Networks

INTRODUCTION

Self-harm and suicide pose significant challenges, particularly in developing countries where 77% of suicide cases occur. This escalating trend is linked to technological advancements and rapid urbanization. Such incidents not only bring personal tragedies but also strain economies due to reduced labor productivity. Traditional methods of tracking these trends through administrative reports are resource-intensive and often yield delayed and coarse-grained data, limiting proactive policymaking. Existing forecasting techniques, like ARIMA and Holt-Winters, have proven inadequate, emphasizing the need for alternative data sources capturing real-time population reactions. This study introduces FAST, a pioneering framework leveraging social media data to forecast aggregate-level self-harm trends. By extracting mental signals from platforms like Twitter and Facebook, FAST combines this data with historical statistics to develop machine-learning-based forecasting models, offering timely and accurate insights for policymakers and healthcare stakeholders.

LITERATURE SURVEY

M. Akyuz and C. Karul*et al*

Author explores the influence of industrial production (IP), inflation, and investment on suicide mortality in Turkey from 1988-2018. Using Fourier cointegration tests and dynamic ordinary least square regression, the study reveals that IP and investment inversely correlate with suicide mortality, indicating a protective effect, while inflation shows a positive correlation, suggesting a detrimental impact. The findings underscore the significance of economic policies like investment promotion, IP enhancement, and disinflation not only for economic growth but also for mental health. This research highlights the need for holistic suicide prevention policies in developing countries, considering broader economic factors beyond just unemployment and GDP.

A. Aldayel and W. Magdy*et al*

Since the author delves into the evolving field of stance detection on social media, a nuanced approach to opinion mining that surpasses traditional sentiment analysis. This paper offers a comprehensive survey across diverse disciplines like natural language processing, web science, and social computing, examining their unique perspectives on stance detection. It presents an in-depth analysis of various techniques, including task definitions, target types, feature sets, and machine learning methods. Highlighting state-of-the-art results and effective strategies, the study also explores emerging trends and applications, such as opinion mining, prediction, and fake news detection. Ultimately, it identifies research gaps and suggests future directions to advance stance detection on social media.

B. E. Belsher, D. J. Smolenski*et al*

As the author underscores the potential of suicide prediction models in enhancing the identification of individuals at elevated suicide risk through predictive algorithms applied to extensive datasets. These models are being tailored for prominent healthcare systems like the US Department of Defense, US Department of Veterans Affairs, and Kaiser Permanente. The study aims to assess the diagnostic accuracy of these models and simulate their impact on population-level suicide rates. After a thorough literature review spanning multiple databases and evaluating 64 unique prediction models across five countries with over 14 million participants, the findings indicate high global classification accuracy but extremely low predictive validity for actual suicide events. Consequently, despite promising classification accuracy, these models currently lack the precision needed for real-world clinical applications.

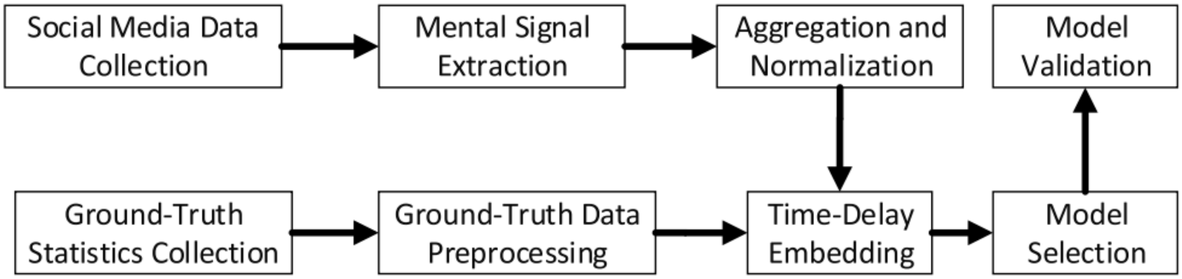
PROBLEM STATEMENT:

Now-a-days world is facing a new disease called depression which is the source to cause all diseases even suicides. Peoples may feel depression due to over competition at education and professional levels. Peoples often express depression from their faces or in their writing post skills. Addressing this problem requires monitoring and forecasting self-harm trends using real-time data from social media, providing timely insights for policy-makers to prevent and mitigate these tragic occurrences.

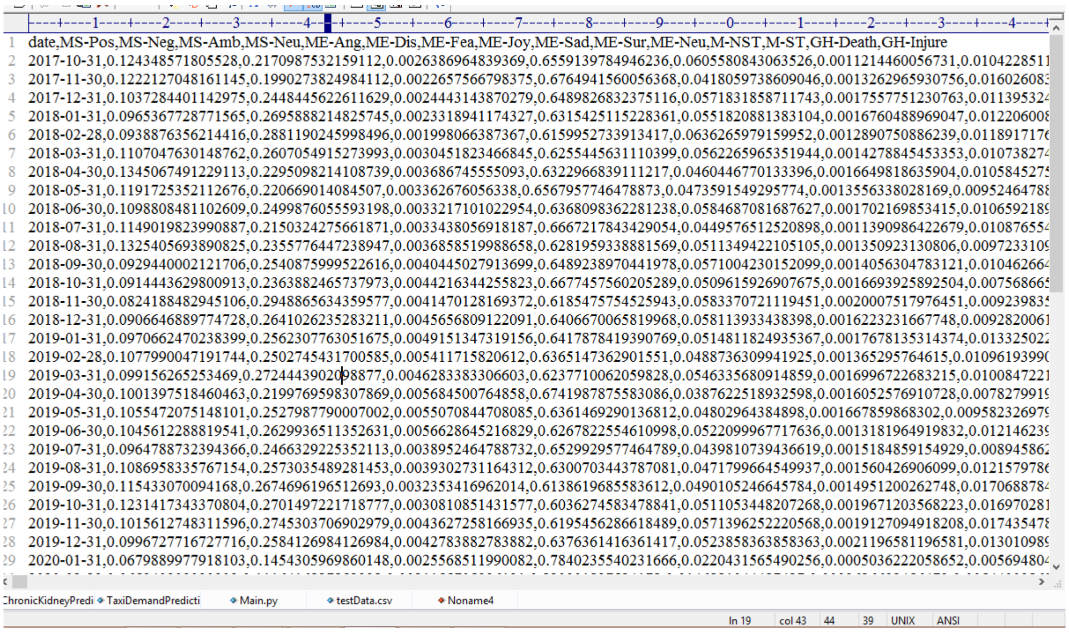
PROPOSED METHOD:

This paper employing FAST (forecast self-harm patterns) novel technique which analyse users post from online social networking sites, then extract emotions and sentiments from tweets to create training dataset. Author has collected suicides and death data from Thailand health department and then combines both emotions and death values to form a dataset. This trained model can be applied on user's new tweet data to forecast self harm activity like injury and death. This prediction can help government to send all those depressed peoples for counselling to reduce self harm activities.

ARCHITECTURE


Social Media Data Collection
Mental Signal Extraction
Aggregation and Normalization
Model Validation
Ground-Truth Statistics Collection
Ground-Truth Data Preprocessing
Time-Delay Embedding
Model Selection

SELF HARM DATASET:



```
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
2017-10-31,0.124348571805528,0.2170987532159112,0.0026386964839369,0.6559139784946236,0.0605580843063526,0.0011214460056731,0.010422851
2017-11-30,0.1222127048161145,0.1990273824984112,0.0022657566798375,0.6764941560056368,0.0418059738609046,0.0013262965930756,0.01602608
2017-12-31,0.1037284401142975,0.2448445622611629,0.0024443143870279,0.6489826832375116,0.0571831858711743,0.0017557751230763,0.01139532
2018-01-31,0.0965367728771565,0.2695888214825745,0.0023318941174327,0.6315425115228361,0.0551820881383104,0.0016760488969047,0.01220600
2018-02-28,0.0938876356214416,0.2881190245998496,0.001998066387367,0.6159952733913417,0.0636265979159952,0.0012890750886239,0.01189171
2018-03-31,0.1107047630148762,0.2607054915273993,0.0030451823466845,0.6255445631110399,0.0562265965351944,0.0014278845453353,0.01073827
2018-04-30,0.1345067491229113,0.2295098214108739,0.003686745555093,0.6322966839111217,0.0460446770133396,0.0016649818635904,0.01058452
2018-05-31,0.1191725352112676,0.220669014084507,0.003362676056338,0.6567957746478873,0.0473591549295774,0.0013556338028169,0.00952464
2018-06-30,0.1098808481102609,0.2499876055593198,0.0033217101022954,0.6368098362281238,0.0584687081687627,0.001702169853415,0.01065921
2018-07-31,0.1149019823990887,0.2150324275661871,0.0033438056918187,0.6667217843429054,0.0449576512520898,0.0011390986422679,0.01087655
2018-08-31,0.1325405693890825,0.2355776447238947,0.0036858519988658,0.6281959338881569,0.0511349422105105,0.001350923130806,0.00972331
2018-09-30,0.0929440002121706,0.2540875999522616,0.0040445027913699,0.6489238970441978,0.0571004230152099,0.0014056304783121,0.01046266
2018-10-31,0.0914443629800913,0.2363882465737973,0.0044216344255823,0.6677457560205289,0.0509615926907675,0.0016693925892504,0.00756866
2018-11-30,0.0824188482945106,0.2948865634359577,0.0041470128169372,0.6185475754525943,0.0583370721119451,0.0020007517976451,0.00923983
2018-12-31,0.0906646889774728,0.2641026235283211,0.0045656809122091,0.6406670065819968,0.058113933438398,0.0016223231667748,0.0092820061
2019-01-31,0.0970662470238399,0.2562307763051675,0.0049151347319156,0.6417878419390769,0.0514811824935367,0.0017678135314374,0.01332502
2019-02-28,0.1077990047191744,0.2502745431700585,0.005411715820612,0.6365147362901551,0.0488736309941925,0.001365295764615,0.01096193
2019-03-31,0.099156265253469,0.2724443902098877,0.0046283383306603,0.6237710062059828,0.0546335680914859,0.0016996722683215,0.01008472
2019-04-30,0.1001397518460463,0.2199769598307869,0.005684500764858,0.6741987875583086,0.0387622518932598,0.0016052576910728,0.00782799
2019-05-31,0.1055472075148101,0.2527987790007002,0.0055070844708085,0.6361469290136812,0.04802964384898,0.001667859868302,0.00958232
2019-06-30,0.1045612288819541,0.2629936511352631,0.0056628645216829,0.6267822554610998,0.0522099967717636,0.0013181964919832,0.01214623
2019-07-31,0.0964788732394366,0.2466329225352113,0.0038952464788732,0.6529929577464789,0.0439810739436619,0.0015184859154929,0.00894586
2019-08-31,0.1086958335767154,0.2573035489281453,0.0039302731164312,0.6300703443787081,0.0471799664549937,0.001560426906099,0.01215797
2019-09-30,0.115433070094168,0.2674696196512693,0.0032353416962014,0.6138619685583612,0.0490105246645784,0.0014951200262748,0.01706887
2019-10-31,0.1231417343370804,0.2701497221718777,0.0030810851431577,0.6036274583478841,0.0511053448207268,0.0019671203568223,0.0169702
2019-11-30,0.1015612748311596,0.2745303706902979,0.0043627258166935,0.6195456286618489,0.0571396252220568,0.0019127094918208,0.01743547
2019-12-31,0.0996727716727716,0.2584126984126984,0.0042783882783882,0.6376361416361417,0.0523858363858363,0.0021196581196581,0.01301098
2020-01-31,0.0679889977918103,0.1454305969860148,0.0025568511990082,0.7840235540231666,0.0220431565490256,0.0005036222058652,0.00569480
```

In above dataset screen first row represents dataset column names and remaining rows represents dataset values and each columns contains emotions and sentiments values like positive, negative, neutral etc. So by using above dataset we will train and test all algorithm performance. In propose paper author evaluated each algorithm performance in terms of MAE, RMSE and MAPE metrics. Each metric represents difference between True Test value and predicted value so the lower the difference the better is the algorithm.

METHODOLOGY:

1. Data Preprocessing:

The initial phase of any data science project involves preparing the dataset for analysis and modeling. Here's how we handle this crucial step:

Importing Essential Libraries: We begin by importing essential Python libraries such as pandas for data manipulation, numpy for numerical computations, and scikit-learn for machine learning tasks.

Reading the Dataset: Our data resides in a CSV file, and we read it into a pandas DataFrame for easier manipulation and analysis.

Handling Missing Values: Data often comes with missing values, which can skew our analysis. We implement strategies like imputation or removal to address these missing values, ensuring data integrity.

Feature-Target Split: To train our predictive models, we separate the dataset into feature variables (X) and target variables, which in this case are 'injury' and 'death counts'.

2. Model Selection and Evaluation:

Now that our data is preprocessed, let's move on to selecting and evaluating predictive models:

Algorithm Selection: We opt for a diverse set of algorithms to capture different aspects of our data's patterns. Our selection includes ARIMA for time series forecasting, Bayesian Ridge for regression tasks, SVR, XGBoost, Random Forest, CatBoost, and Decision Trees.

Data Splitting: Before training our models, we split our dataset into training and testing subsets using train_test_split from scikit-learn, ensuring that our models are evaluated on unseen data.

Training and Prediction: We train each selected model on the training data and then make predictions on the test set.

Performance Evaluation: To gauge the efficacy of our models, we employ key performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Visualization: For a more intuitive understanding of our models' predictions, we visualize the predicted values against the actual values using plots or graphs.

3. Algorithm Comparison:

With predictions and evaluations in hand, it's time to compare the performance of our selected algorithms:

Performance Metrics Comparison: We juxtapose the performance of all algorithms based on our chosen evaluation metrics, providing insights into their relative strengths and weaknesses.

Tabular Summary: To simplify the comparison, we summarize the results in a tabular format, detailing the performance of each algorithm for both 'injury' and 'death' predictions.

4. Prediction on Test Data:

Now, let's apply our best-performing model to make predictions on new, unseen data:

Reading Test Data: We read the test data, stored in a CSV file, into a pandas DataFrame, ensuring consistency with our training data preprocessing.

Handling Missing Values: As before, we address any missing values in the test dataset to ensure robust predictions.

Final Predictions: Leveraging the insights from our evaluations, we deploy the best-performing model, in this case, the Decision Tree, to make predictions on the test data for both 'injury' and 'death' counts.

Mean Squared Error (MSE):

MSE is calculated by taking the average of the squared differences between the predicted values and the actual values.

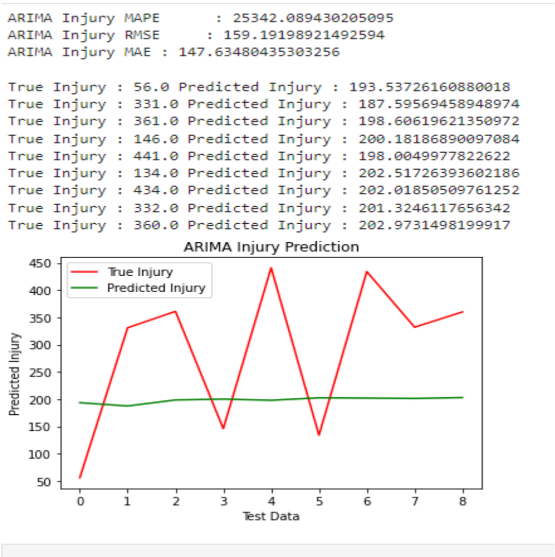
$$\mathrm{MSE} = \frac{1}{n} \quad {}_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Root Mean Squared Error (RMSE):

RMSE is the square root of the MSE and provides a measure of the average magnitude of the errors in the predictions.

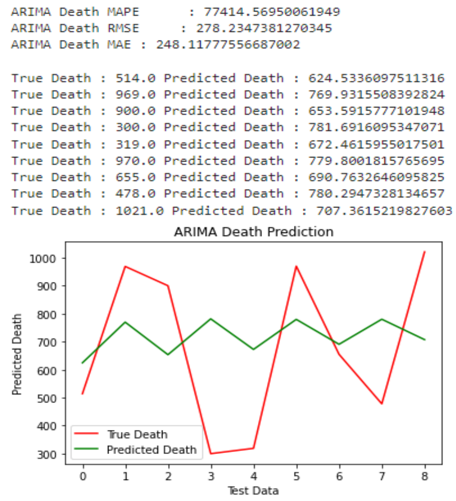
$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$$

RESULTS:



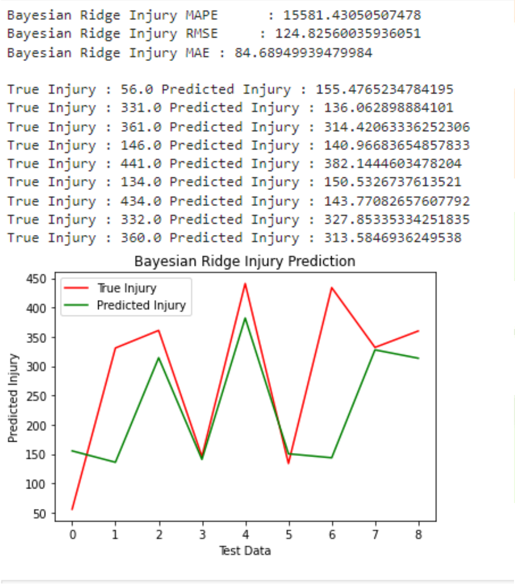
```
ARIMA Injury MAPE      : 25342.089430205095
ARIMA Injury RMSE     : 159.19198921492594
ARIMA Injury MAE : 147.63480435303256

True Injury : 56.0 Predicted Injury : 193.53726160880018
True Injury : 331.0 Predicted Injury : 187.59569458948974
True Injury : 361.0 Predicted Injury : 198.60619621350972
True Injury : 146.0 Predicted Injury : 200.18186890097084
True Injury : 441.0 Predicted Injury : 198.0049977822622
True Injury : 134.0 Predicted Injury : 202.51726393602186
True Injury : 434.0 Predicted Injury : 202.01850509761252
True Injury : 332.0 Predicted Injury : 201.3246117656342
True Injury : 360.0 Predicted Injury : 202.9731498199917
```

In above graph x-axis represents number of test data and y-axis represents Injury values and red line represents True Injury and green line represents Predicted injury and there is lots of gaps between red and green .line so ARIMA performance is not good. If predicted values are accurate then both lines will overlap.



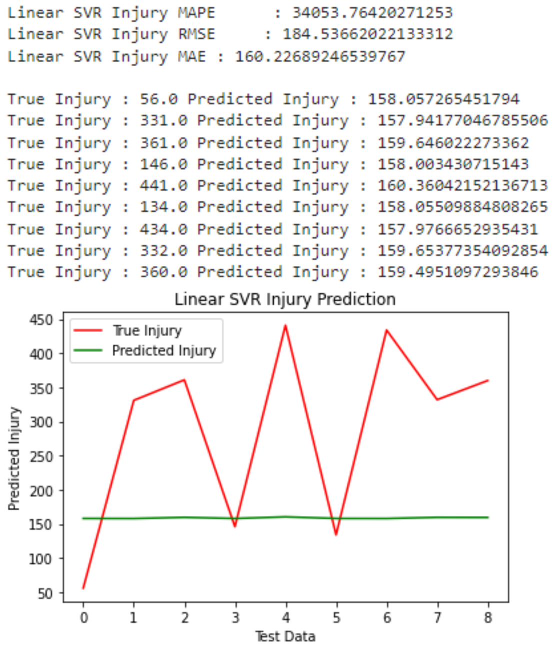
```
ARIMA Death MAPE      : 77414.56950061949
ARIMA Death RMSE     : 278.2347381270345
ARIMA Death MAE : 248.11777556687002

True Death : 514.0 Predicted Death : 624.5336097511316
True Death : 969.0 Predicted Death : 769.9315508392824
True Death : 900.0 Predicted Death : 653.5915777101948
True Death : 300.0 Predicted Death : 781.6916095347071
True Death : 319.0 Predicted Death : 672.4615955017501
True Death : 970.0 Predicted Death : 779.8001815765695
True Death : 655.0 Predicted Death : 690.7632646095825
True Death : 478.0 Predicted Death : 780.2947328134657
True Death : 1021.0 Predicted Death : 707.3615219827603
```

In above screen training Bayesian Ridge on Injury data and then displaying its performance MAE and other values and in graph both lines are overlapping with little gap



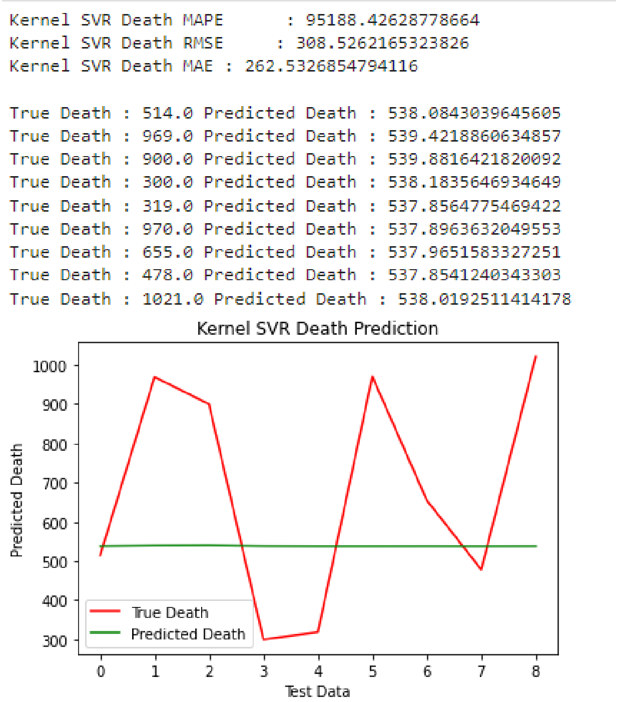
```
Bayesian Ridge Injury MAPE      : 15581.43050507478
Bayesian Ridge Injury RMSE     : 124.82560035936051
Bayesian Ridge Injury MAE : 84.68949939479984

True Injury : 56.0 Predicted Injury : 155.4765234784195
True Injury : 331.0 Predicted Injury : 136.062898884101
True Injury : 361.0 Predicted Injury : 314.42063336252306
True Injury : 146.0 Predicted Injury : 140.96683654857833
True Injury : 441.0 Predicted Injury : 382.1444603478204
True Injury : 134.0 Predicted Injury : 150.5326737613521
True Injury : 434.0 Predicted Injury : 143.77082657607792
True Injury : 332.0 Predicted Injury : 327.85335334251835
True Injury : 360.0 Predicted Injury : 313.5846936249538
```

Above is the Bayesian ridge training on death data



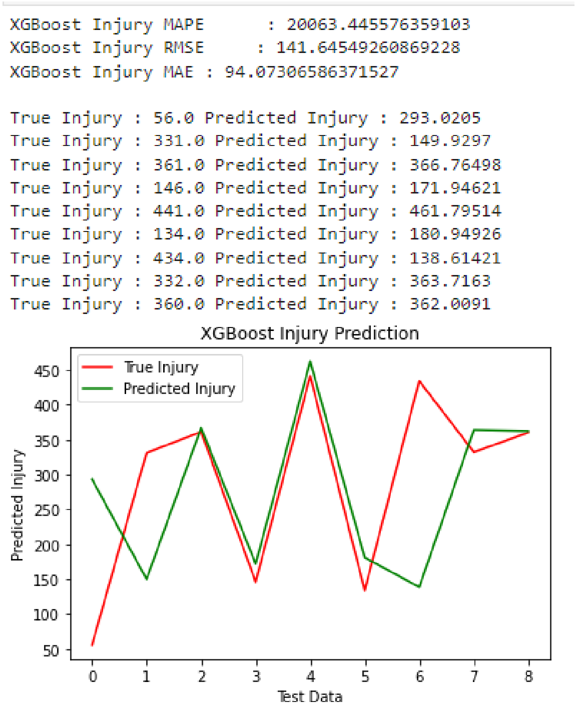
```
Linear SVR Injury MAPE      : 34053.76420271253
Linear SVR Injury RMSE     : 184.53662022133312
Linear SVR Injury MAE : 160.22689246539767

True Injury : 56.0 Predicted Injury : 158.057265451794
True Injury : 331.0 Predicted Injury : 157.94177046785506
True Injury : 361.0 Predicted Injury : 159.646022273362
True Injury : 146.0 Predicted Injury : 158.003430715143
True Injury : 441.0 Predicted Injury : 160.36042152136713
True Injury : 134.0 Predicted Injury : 158.05509884808265
True Injury : 434.0 Predicted Injury : 157.9766652935431
True Injury : 332.0 Predicted Injury : 159.65377354092854
True Injury : 360.0 Predicted Injury : 159.4951097293846
```

Above is the SVM training on Injury data



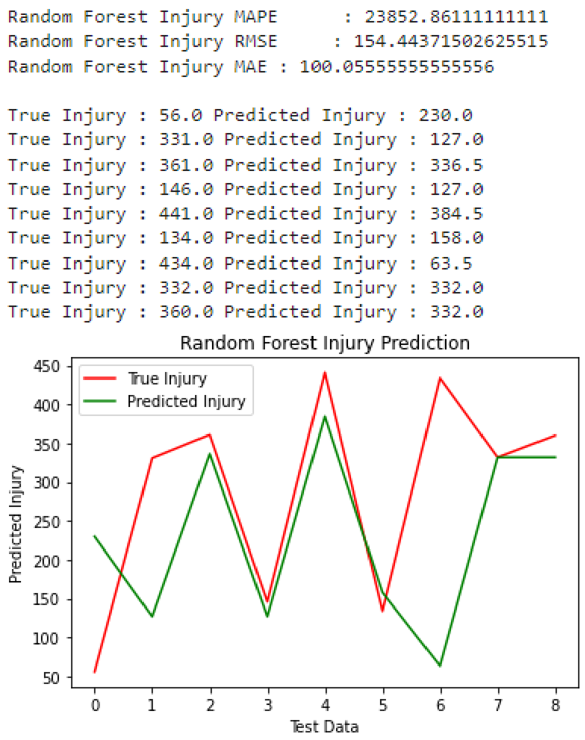
```
Kernel SVR Death MAPE      : 95188.42628778664
Kernel SVR Death RMSE     : 308.5262165323826
Kernel SVR Death MAE : 262.5326854794116

True Death : 514.0 Predicted Death : 538.0843039645605
True Death : 969.0 Predicted Death : 539.4218860634857
True Death : 900.0 Predicted Death : 539.8816421820092
True Death : 300.0 Predicted Death : 538.1835646934649
True Death : 319.0 Predicted Death : 537.8564775469422
True Death : 970.0 Predicted Death : 537.8963632049553
True Death : 655.0 Predicted Death : 537.9651583327251
True Death : 478.0 Predicted Death : 537.8541240343303
True Death : 1021.0 Predicted Death : 538.0192511414178
```

Above is the SVM training on death data



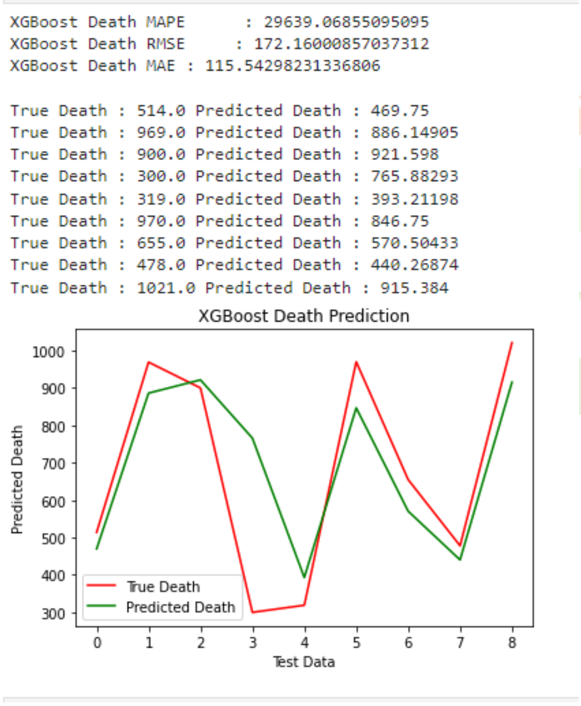
```
XGBoost Injury MAPE      : 20063.445576359103
XGBoost Injury RMSE     : 141.64549260869228
XGBoost Injury MAE : 94.07306586371527

True Injury : 56.0 Predicted Injury : 293.0205
True Injury : 331.0 Predicted Injury : 149.9297
True Injury : 361.0 Predicted Injury : 366.76498
True Injury : 146.0 Predicted Injury : 171.94621
True Injury : 441.0 Predicted Injury : 461.79514
True Injury : 134.0 Predicted Injury : 180.94926
True Injury : 434.0 Predicted Injury : 138.61421
True Injury : 332.0 Predicted Injury : 363.7163
True Injury : 360.0 Predicted Injury : 362.0091
```

Above is the XGBOOST training on injury data



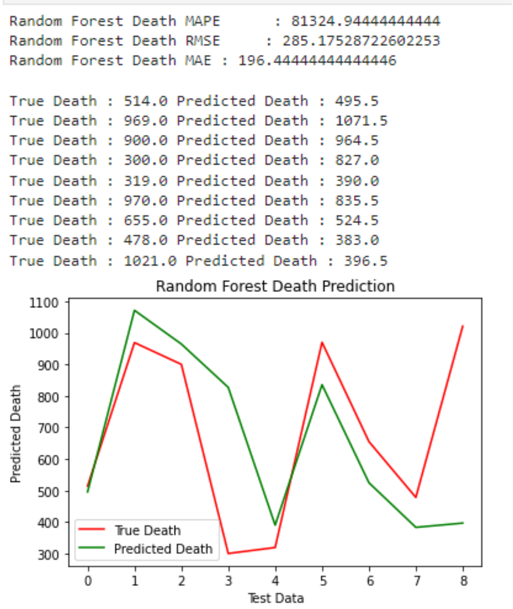
```
Random Forest Injury MAPE      : 23852.86111111111
Random Forest Injury RMSE     : 154.44371502625515
Random Forest Injury MAE : 100.05555555555556

True Injury : 56.0 Predicted Injury : 230.0
True Injury : 331.0 Predicted Injury : 127.0
True Injury : 361.0 Predicted Injury : 336.5
True Injury : 146.0 Predicted Injury : 127.0
True Injury : 441.0 Predicted Injury : 384.5
True Injury : 134.0 Predicted Injury : 158.0
True Injury : 434.0 Predicted Injury : 63.5
True Injury : 332.0 Predicted Injury : 332.0
True Injury : 360.0 Predicted Injury : 332.0
```

Above is the random forest training on Injury data



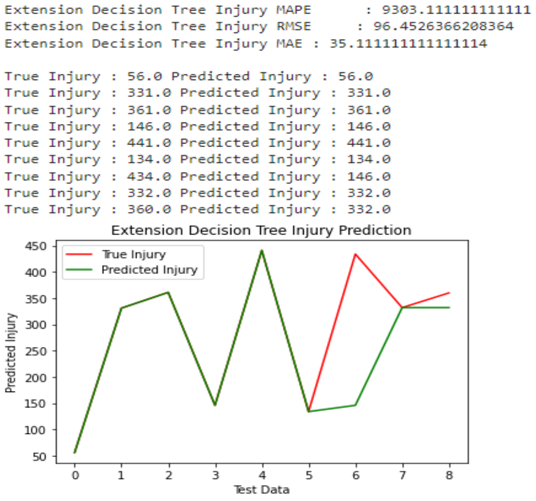
```
XGBoost Death MAPE      : 29639.06855095095
XGBoost Death RMSE     : 172.16000857037312
XGBoost Death MAE : 115.54298231336806

True Death : 514.0 Predicted Death : 469.75
True Death : 969.0 Predicted Death : 886.14905
True Death : 900.0 Predicted Death : 921.598
True Death : 300.0 Predicted Death : 765.88293
True Death : 319.0 Predicted Death : 393.21198
True Death : 970.0 Predicted Death : 846.75
True Death : 655.0 Predicted Death : 570.50433
True Death : 478.0 Predicted Death : 440.26874
True Death : 1021.0 Predicted Death : 915.384
```

Above is the XGBOOST training on Death data



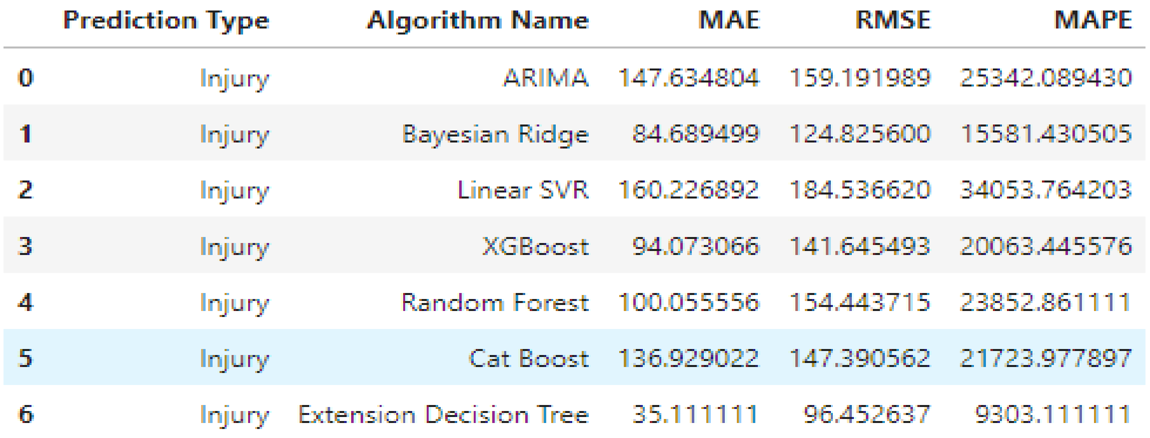
```
Random Forest Death MAPE      : 81324.94444444444
Random Forest Death RMSE     : 285.17528722602253
Random Forest Death MAE : 196.44444444444446

True Death : 514.0 Predicted Death : 495.5
True Death : 969.0 Predicted Death : 1071.5
True Death : 900.0 Predicted Death : 964.5
True Death : 300.0 Predicted Death : 827.0
True Death : 319.0 Predicted Death : 390.0
True Death : 970.0 Predicted Death : 835.5
True Death : 655.0 Predicted Death : 524.5
True Death : 478.0 Predicted Death : 383.0
True Death : 1021.0 Predicted Death : 396.5
```

Above is the random forest training on Death data



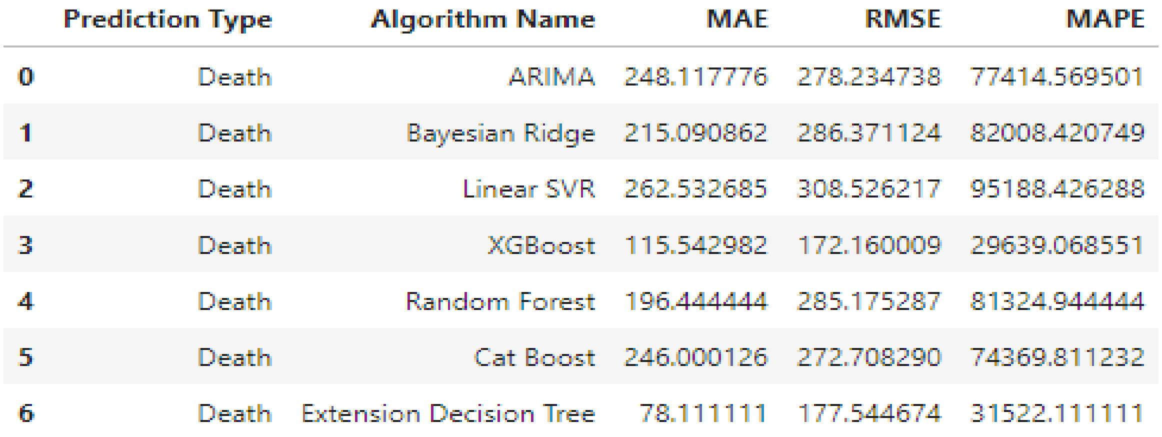
```
Extension Decision Tree Injury MAPE      : 9303.111111111111
Extension Decision Tree Injury RMSE     : 96.4526366208364
Extension Decision Tree Injury MAE : 35.111111111111114

True Injury : 56.0 Predicted Injury : 56.0
True Injury : 331.0 Predicted Injury : 331.0
True Injury : 361.0 Predicted Injury : 361.0
True Injury : 146.0 Predicted Injury : 146.0
True Injury : 441.0 Predicted Injury : 441.0
True Injury : 134.0 Predicted Injury : 134.0
True Injury : 434.0 Predicted Injury : 146.0
True Injury : 332.0 Predicted Injury : 332.0
True Injury : 360.0 Predicted Injury : 332.0
```

In above screen training extension decision tree on Death data and can see its performance metrics



	Prediction Type	Algorithm Name	MAE	RMSE	MAPE
0	Injury	ARIMA	147.634804	159.191989	25342.089430
1	Injury	Bayesian Ridge	84.689499	124.825600	15581.430505
2	Injury	Linear SVR	160.226892	184.536620	34053.764203
3	Injury	XGBoost	94.073066	141.645493	20063.445576
4	Injury	Random Forest	100.055556	154.443715	23852.861111
5	Injury	Cat Boost	136.929022	147.390562	21723.977897
6	Injury	Extension Decision Tree	35.111111	96.452637	9303.111111

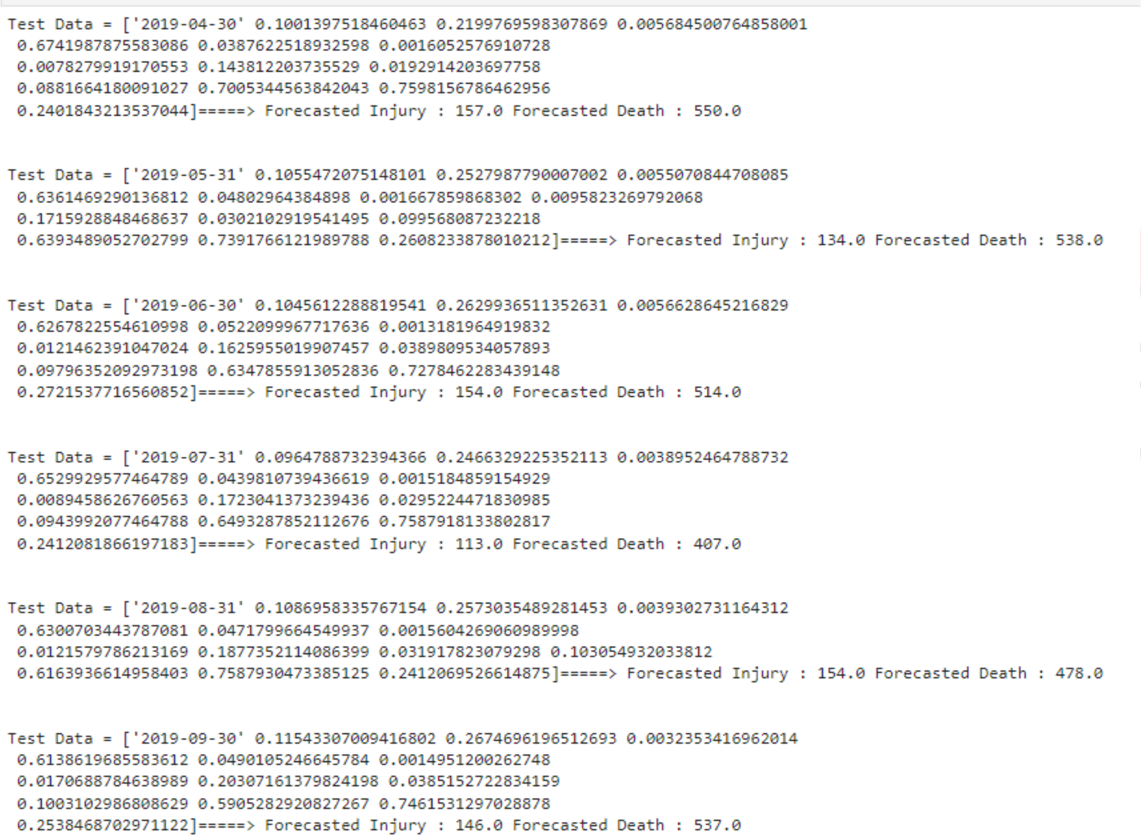
In above screen can see each algorithm MAE, RMSE and MAPE error values of all algorithms on INJURY data and this metrics are represented as error so if error values are less then algorithm will be consider as best and in all algorithms XGBOOST and extension decision tree got less error rates



	Prediction Type	Algorithm Name	MAE	RMSE	MAPE
0	Death	ARIMA	248.117776	278.234738	77414.569501
1	Death	Bayesian Ridge	215.090862	286.371124	82008.420749
2	Death	Linear SVR	262.532685	308.526217	95188.426288
3	Death	XGBoost	115.542982	172.160009	29639.068551
4	Death	Random Forest	196.444444	285.175287	81324.944444
5	Death	Cat Boost	246.000126	272.708290	74369.811232
6	Death	Extension Decision Tree	78.111111	177.544674	31522.111111

In above screen can see each algorithm MAE, RMSE and MAPE error values of all algorithms on DEATH data and in all algorithms XGBOOST and extension decision tree got less error rates

PREDICTION:



```
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


Test Data = ['2019-07-31' 0.0964788732394366 0.2466329225352113 0.0038952464788732
 0.6529929577464789 0.0439810739436619 0.0015184859154929
 0.0089458626760563 0.1723041373239436 0.0295224471830985
 0.0943992077464788 0.6493287852112676 0.7587918133802817
 0.2412081866197183]=====> Forecasted Injury : 113.0 Forecasted Death : 407.0


Test Data = ['2019-08-31' 0.1086958335767154 0.2573035489281453 0.0039302731164312
 0.6300703443787081 0.0471799664549937 0.0015604269060989998
 0.0121579786213169 0.1877352114086399 0.031917823079298 0.103054932033812
 0.6163936614958403 0.7587930473385125 0.2412069526614875]=====> Forecasted Injury : 154.0 Forecasted Death : 478.0


Test Data = ['2019-09-30' 0.11543307009416802 0.2674696196512693 0.0032353416962014
 0.6138619685583612 0.0490105246645784 0.0014951200262748
 0.0170688784638989 0.20307161379824198 0.0385152722834159
 0.1003102986808629 0.5905282920827267 0.7461531297028878
 0.2538468702971122]=====> Forecasted Injury : 146.0 Forecasted Death : 537.0
```

In above screen reading test dataset values and then performing forecasting of INJURY and DEATH and in output before arrow symbol we can see test data values and after arrow symbol can see forecasted values of INJURY and DEATH

CONCLUSION

Depression is a growing concern worldwide, often leading to self-harm and even suicides. This project aims to forecast self-harm trends using social networks, leveraging the FAST (Forecast Self-Harm Patterns) technique. By analyzing users' online posts, emotions, and sentiments, we trained various machine learning algorithms to predict injury and death occurrences. Notably, XGBoost and the extension Decision Tree algorithm outperformed others in forecasting accuracy. The generated dataset and models hold promise for early intervention strategies, guiding governments in providing timely counseling and support to mitigate self-harm activities.

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.