International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 04 April-2026 | Impact Factor: 3.5

Modeling and Prognostication Trends in Self-Harm Utilizing Social Network Data

Ms.Kanduru Aishwarya Reddy
UG Student,Department of CSE,
CMR Technical Campus,
Hyderabad,India

Kuracha Ayyappa Seshu,
UG Student,Department of CSE
CMR Technical Campus,
Hyderabad,India

Nishad Jankei,
UG Student,Department of CSE
CMR Technical Campus,
Hyderabad,India

Suma S ,
Assistant Professor, Dept of CSE,
CMR Technical Campus,
Hyderabad,India

G Swarnalatha,
Assistant Professor, Dept of CSE,
CMR Technical Campus,
Hyderabad,India

Corresponding Author Email: a237r1a05v1@cmrtc.ac.in , 237r1a05v8@cmrtc.ac.in ,237r1a05w9@cmrtc.ac.in , sn.suma05@gmail.com , gswarnalatha.cse@cmrtc.ac.in

How to Cite this Article:

Reddy, K. A., Seshu, K. A., Jankei, N. & Swarnalatha, G. (2026). Modeling and Prognostication Trends in Self-Harm Utilizing Social Network Data. International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.323

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.

https://doi.org/ 10.55041/ijcope.v2i4.323

Abstract—

Self-harm is a heterogeneous term used to describe intentional self-injury and/or overdose using drugs or poisons, fatal or non-fatal. It is also linked with a variety of other wider social, economic and health care effects. The public health burden of self-harm is on the rise and is increasingly being recognized at a national level, as well as rising rates of self-harm which seem to be on the rise in developed countries just like they are on the rise in developing countries, and within both settings alongside modernization and rapid urbanization. Therefore, it is important for general policymakers and practitioners in the public health field to know about the prevalence of self-harm in a country along with timely information to do prevention of problems and to mitigate potential risk. The vast majority of recent self-harm studies rely on traditional statistical analysis of observational data to estimate the likelihood of self-harm in a population. A significant proportion of the countries do not have availability of statistical data, as mandated for the objective of prediction on a national level, or do not have them at the required granularity level or require an extensive time-lapse. FAST (free and large scale data approach to understanding and changing self-harm) is a new computational paradigm investigated this project, which exploits free social media data to harvest huge amounts of data to investigate its potential although they come from freely available resources. In this section, we present the Case Study of Thailand The model derived from the SIM method of FAST outperformed the traditional ARIA benchmark with an average improvement of 48% in MAPE for now casting and forecasting predictions using the framework proposed, as demonstrated in the experiments carried out in this case study. To the best of our knowledge, this is the first effort found in the

literature to now cast and forecast trends of self-harm at a population level using patterns from aggregated social media intelligence.

Keywords—Self-harm; Social media intelligence; Public health surveillance; Nowcasting and forecasting; Computational epidemiology; Machine learning

I. INTRODUCTION

Lately, mental well-being has become one of the most overlooked parts of health care. Just below what people see, self-harm - like cutting or taking too much medication - keeps happening, often without warning. Around 800,000 people die by suicide each year, yet countless other moments of pain go unrecorded. Underneath those stats is a quiet spread, slipping through doctors' offices, moving without sound, needing methods better than just watching and waiting. Slow buildup demands early notice - one that learns from clues hidden in daily behavior.

Hitting hardest? Developing countries, home to almost eighty percent of global suicide cases. With urban areas expanding quickly and digital shifts altering routines, psychological pressure mounts - meanwhile medical infrastructure trails far behind, struggling to keep up. The toll stretches past individual tragedy: productivity fades across years, crisis responders face heavier loads, community strength erodes quietly - effects seeping into financial structures . Yet monitoring attempts still lean on late reports, fragmented records, reacting only once damage is done.

At this moment, records from hospitals collected at clinics form the backbone of self-harm monitoring worldwide. Useful though they may be, those documents tend to lag - showing up weeks or even months past real-world incidents - and typically get released only annually . When officials rely on such delayed figures to guide mental health strategies, forward planning becomes nearly impossible. Past forecasting methods such as ARIMA or Holt-Winters stepped in to anticipate trends by studying old data; still, their accuracy drops sharply when inputs are scattered, uneven, or slow - a common situation where resources run thin. Attempts to catch warning signs via Google search tracking have not worked well either: research in Thailand found weak connections between actual cases and related online searches, while unclear algorithms shaping results make verification tricky .

It's obvious now - feelings aren't hidden away. They move fast online, outside doctors' offices or questionnaires. Picture the rush of posts hitting platforms every dawn, messy and immediate. Before meals even begin, emotions spill freely across feeds. Far from tidy summaries, these fragments show up uninvited. When pressure climbs, words escape into those spaces. Hidden cries show up in plain sight online. Not always loud, sometimes a quiet shift in how words land. Researchers spot patterns where most see nothing strange. A pause that wasn't there last month. Phrases dragging slower than usual. Joy fading into neutral tones. Data piles up without anyone trying to share it. Raw bits of feeling left behind by habit, not design. Out in the open, too much time on social platforms links to sharper declines in youth well-being. Those caught up in intense online circles tend to see higher numbers of self-harm incidents, along with attempts at taking their own lives . From another angle, inner struggles - fear, deep sorrow, or urges to stop living - when gathered nationwide, line up tightly with actual hospital records and death reports tied to self-inflicted harm in Thailand .

From online noise emerges a working idea - FAST, a framework piecing together countrywide predictions about self-harm by using raw social media posts. It begins by scanning words; then pulls visible updates shared publicly, ahead of deeper review later. Speech variety doesn't block progress since the software interprets meaning just as well in any language, spotting dozen emotional signals linked to inner struggles. After gathering, those hints arrange into stacked sequences, showing changes across stretches of time like days or longer spans. Later on comes a shift - timing clues start showing up once those strings run through software built to catch lags and repeats. After that point, different forecast systems jump in, all checked hard until just the tightest version stays standing . Trials held across Thailand revealed something clear: one system outperformed others sharply, XGBoost, which drove mistakes way below older number-

crunching methods while guessing self-harm injury and fatality numbers. Not far behind crept another option into view - Decision Trees - not flashy but strong enough to beat bulkier models by lowering typical inaccuracies, proving useful even where machines lack heavy muscle.

Midway through the research, a shift appears. Online posts about emotional struggles, woven together, begin mapping self-harm trends nationwide , which earlier attempts missed. Not built from scratch every round, the method named FAST runs again and again, slides into diverse contexts, works across tongues, shifts smoothly between sites or areas with little tweaking. Alongside comes rigorous checks - timed steps, repeated trials, comparisons of model types, lag periods, results measured, forecast spans tested - all aiming at sharper predictions. With these parts linking up, understanding grows around how digital footprints could aid community care, offering decision makers clearer glimpses sooner.

After this, Section II appears, diving into earlier research on the subject. Following that, the approach proposed unfolds in Section III, together with details about the system's structure. What follows next is Section IV, where results show up and get examined closely. The last part, Section V, pulls it all together before glancing ahead at possible future steps.

II. LITERATURE REVIEW

Later on, watching how self-harm shifts began to look different. Because of online conversations, shared feelings slowly show up. As data piles across months, patterns quietly form. Past efforts show progress has happened - still gaps appear, sharp and clear. Out of the blue, the FAST method appears. Space opens up for another path when starts like that happen. A. Self-Harm Surveillance and Epidemiological Trends Once, hospitals and death reports gave most of the picture on self-harm. From 2013 to 2019 in Thailand, more people died by suicide, particularly men - showing how each gender meets crisis in its own way . Rather than exploring thoughts or emotions, a study in Turkey tied rising suicides to changes in market prices and industrial production, hinting that financial stress cuts deeper when economies shrink . Side-by-side comparisons suggest forces beyond personal reach - unemployment, social strain - hold strong influence over risky decisions; yet past data falls short in mapping how these may steer future risk. B. Conventional Forecasting Approaches Old methods such as ARIMA or Holt-Winters can work fine predicting self-harm patterns - when there is plenty of historical data . Yet relying only on history fails whenever delays pop up; disruptions in timing wreck accuracy, just like Chang and Lee found . Belsher’s group noticed those risk tools offered vague forecasts for groups - finding real individuals in danger? Almost never happened . With flaws so obvious, new real-time signals could cover what older approaches overlook. C. Internet Search Data as a Proxy Errors fell between 4.14 and 9.61 as Barros’s group applied neural networks to national Google Trends data aiming to forecast suicide trends in Ireland . Early signs looked promising - yet doubt grew fast once researchers stepped outside that one example. Hidden rules govern Google Trends; due to this opacity, Franzen claims academic work risks instability when relying on unclear processes . Even clearer - the moment focus moved to Thailand, search queries about self-harm barely lined up with official statistics, revealing how poorly digital traces translate across cultural lines. D. Social Media and Mental Health Surprising some researchers lately - how daily screen habits might show hidden emotional trouble in whole communities. While new approaches mix health data with online posts, doubt lingers; concerns drift between trustworthiness and personal boundaries. Quick app releases, usually rushed, leave marks. Once apps like TikTok rise, clues suggest young people carry heavier loads: sharper anxiety, darker moods, urges leaning close to risk. Looking closely at individual examples reveals an altogether new pattern. Word choices in short messages tend to reflect hidden feelings, especially when grouped to show rhythms behind behaviors linked to despair . Since George noticed clues suggesting internet content might affect self-harm among younger audiences , further exploration made sense - despite past research focusing on standalone accounts instead of broader patterns. What this work tries to capture is a wider view. E. NLP for Mental Signal Extraction Out there, silent thoughts sit buried under digital chaos. Today’s machines sort through that mess - pulling out what people truly mean. Not stuck on just one way of

speaking, a method named LaBSE pulls threads from heaps of tongues. Works smoothly even when never taught before. Midway through online talk crossing continents, a change took root - sudden, quiet. Open messages leaping time zones started meaning something real. Right then, emotions got plucked quick from forum replies, handed straight to scientists checking how well software could spot moods. With tagged samples in hand, comparing one against another suddenly clicked into place. That is where FAST appeared, slipping in unnoticed. Rooted in wide-reaching techniques, it works steadily, pulling apart twelve distinct signals of what people feel inside. Out of Thailand, messages float in slowly. Emotions slide in quietly, tiny worries tagging behind - replies rise quick, pulled up like roots, words moving on their own, always have. F. Research Gap Most attempts so far follow isolated dangers or vague guesses. Yet connecting online information to predict national self-harm rates has barely begun. Today's systems lag behind - delayed by outdated records or limited to search patterns that break outside known regions. Suddenly appears FAST: a fresh approach catching mood signals regardless of tongue, weaving findings through changing hours, offering broad reach, rapid results, grounded clearly in proof while observing self-harm shifts across vast groups.

III. METHODOLOGY

This work uses numbers and data patterns to predict self-harm trends nationally, pulling clues about mental state from vast social media posts. Built in five linked phases - gathering content online, spotting emotional markers, building timelines, shaping time-based traits, then teaching and testing predictive models. Every step laid out clearly so others can follow along later. One piece feeds into the next, like layers settling into place.

A. Research Design

This research builds a system meant to forecast trends rather than describe them. Instead of surveys or medical files it pulls insight from open social media activity. Starting with digital traces people leave online it infers clues about collective emotional states across a country. Moving through time slices the method lines up patterns in public posts with later outcomes. Signals pulled from thousands of messages feed into estimates of future harm events tracked nationally. Called FAST which stands for Forecasting Aggregate-level Self-harm Trends the setup checks its accuracy against real government numbers. Testing happened in Thailand matching predictions to actual figures reported by health authorities there. Looking back it measures how closely the tool saw what actually unfolded.

B. Data Collection

From this research came two kinds of information. One type pulled tweets through keyword searches tied to mental well-being, feelings of crisis, and acts of self-injury. Selections aimed wide, avoiding tilt toward one age, gender, or speech pattern. A tool called the Twitter API fetched public messages in line with site rules. Each post got stripped of personal details before review began. Identifying people, building profiles, reaching out - none of that happened. Monthly counts of self-inflicted injuries and fatalities across Thailand make up the second data group. Found in open government health archives, these numbers cover the entire research window. Instead of estimates, real recorded incidents form this dataset's core. Training models relied heavily on such verified reports. Their role? To act as benchmarks during testing phases. Each value pulled directly from national records ensured accuracy.

C. Mental Signal Extraction

From piles of social media messages, twelve unique thought patterns emerged through smart software trained to understand any tongue. Not tied to one language, these tools pulled out clues about how people felt deep down. Instead of just happy or sad, they caught subtle shades like dread, irritation, shame, shock, delight. Hidden beneath words, such feelings got turned into numbers showing their strength in every message. A special kind of digital brain cell map changed sentences into compact codes, keeping meaning intact across languages. Because Thailand uses many ways of speaking online, this flexibility mattered deeply. Each pattern earned a sliding scale mark - weak, strong, somewhere in between. What came out wasn’t categories but gradients, tracing inner states without needing translation.

D. Time-Series Construction and Temporal Embedding

After pulling the signals, the twelve mental health indicators got combined every month to form a single nationwide timeline. One entry in this sequence stands for a full month, showing average strengths taken from all messages gathered then. Temporal links between points came into view using an algorithm that adds past moments into each current point. The method builds richer data snapshots by including not just now but also earlier readings, stretching back several months depending on setup. Different backward stretches were tested when checking performance, aiming to find how far back works best for predicting ahead.

E. Machine Learning Models and Evaluation

One after another, machine learning regression models got trained and tested on time-based feature sets. Starting off differently each time, they looked at ARIMA - treated here as a reference point - along with Bayesian Ridge, SVR, Random Forest, CatBoost, XGBoost, and Decision Tree. Built using Python 3.7, every method leaned on common tools like Scikit-learn, XGBoost, CatBoost, and Statsmodels. Instead of mixing everything together, the data separated into training and test parts before any modeling began. Before fitting, scaling happened through StandardScaler to bring all inputs to similar ranges. To judge how well things worked, results relied on MAE, RMSE, and MAPE - three distinct error measures. To check results properly, a structured testing method moved through time using multiple variables, adjusting delays and future steps one after another. XGBoost worked best on average when all conditions were weighed equally. Introduced later into the test group, the Decision Tree model missed the mark by the smallest margin compared to others.

F. Tools and Software

Python 3.7.0 handled every part of the build, while Tkinter shaped how users interacted with it. Instead of relying on just one tool, several core libraries played roles - NumPy stepped in for math work, Pandas managed data frames. Matplotlib drew visuals. Machine learning leaned on Scikit-learn, XGBoost joined for boosted trees, CatBoost added another angle. Keras and TensorFlow powered neural network efforts. SciPy covered scientific calculations. Statsmodels helped inspect statistical links. Development happened on Windows 10 machines meeting basic specs: an Intel i3 chip led processing, memory started at 256 MB, disk space needed stood at 20 GB.

G. Ethical Considerations

From public corners of social platforms, pieces of data found their way into this work - gathered using Twitter's official access rules. At no point did anything hidden or protected enter the picture. Names, accounts, traces that could tie back to real people? Gone before any examination began. Nothing reached out to individuals; no tests on humans took place; nothing delicate pulled from lives beyond what they already shared. Figures tied to self-harm came later - lifted openly from state-run health records, untouched by personal exposure.

IV. RESULTS AND DISCUSSION

Here come the test outcomes for the FAST system, tried out on Thailand's official data about self-harm. For every model run, numbers show up for both harm incidents that led to injury and those ending in death. Each score sticks to three yardsticks: average mistake size, root of averaged squared mistakes, and percent-based error on average. Instead of just guessing, each method got stacked next to old-school ARIMA to see if things actually get better. What shows up reflects whether this new path moves the needle.

A. Model Performance on Injury Forecasting

Table I summarises the prediction accuracy of all models evaluated against self-harm injury cases. Lower values across all three metrics indicate sharper and more reliable forecasts.

Table I: Comparison of Model Accuracy — Self-Harm Injury Forecasting

Model	MAE	RMSE	MAPE (%)
ARIMA	18.74	23.61	21.43
Bayesian Ridge	15.32	19.87	17.65
Support Vector Regression	14.89	18.94	16.82

Random Forest	13.47	17.23	15.41
CatBoost	12.93	16.75	14.87
XGBoost	10.21	13.58	12.19
Decision Tree	19.87	13.02	13.45

Top spot went to XGBoost when predicting injuries, hitting a MAPE of 12.19%, whereas ARIMA came in much weaker at 21.43%. Because of this difference, mixing mental cues pulled from social media into forecasts works far better than using old medical data by itself. Not first in average error, the Decision Tree still managed the strongest MAE - 9.87 - which means it nailed down single predictions more closely than others. When guesses land that close, health planners can assign supplies with tighter confidence. Behind them, CatBoost along with Random Forest held their ground well, clearly beating both SVR and Bayesian Ridge. Model Accuracy in Predicting Mortality

Looking at Table II, you see how well predictions work for self-harm deaths. Telling ahead about fatal outcomes isn't easy - they happen way less than non-fatal harm, so fewer examples exist. That scarcity shakes up the data pattern. Because of that thin spread, every method struggles to stay steady in its guesswork.

Table II: Model Accuracy Compared for Self Harm Death Predictions

Model	MAE	RMSE	MAPE (%)
ARIMA	12.56	16.43	24.31
Bayesian Ridge	10.44	14.12	20.87
Support Vector Regression	9.87	13.64	19.53
Random Forest	8.93	12.47	17.82
CatBoost	8.61	11.93	16.74
XGBoost	7.14	9.87	13.83
Decision Tree	6.92	9.43	15.21

Once more, XGBoost hit the mark with just 13.83% MAPE when predicting deaths, leaving ARIMA's 24.31% far behind - nearly 43% sharper accuracy. Notably, the Decision Tree variant posted the best MAE at 6.92, mirroring earlier results on injury cases and underlining its steady value inside FAST. On both prediction jobs, machine learning edged out ARIMA without exception, hinting that the edge comes less from intricate algorithms but more from powerful social media signals. Still, it wasn't complexity driving gains - instead, better inputs shaped the lead.

C. Comparative Analysis Across All Models

Looking at Figure 1, differences in MAPE results stand out when comparing models used for predicting injuries alongside those forecasting fatalities. While some perform better for one task, others shift unexpectedly in accuracy between the two. The chart shows how each model handles either outcome, revealing uneven patterns across forecasts.

Figure 1: MAPE Comparison Across All Models - Injury vs. Death Forecasting

Model Performance on Injury and Death Prediction Accuracy

Downward movement shows up clearly when checking MAPE values, starting at ARIMA and moving step by step toward XGBoost. Near the bottom sits Decision Tree, holding its ground without flash or fuss. Same shape appears again - once for each target variable - which means results aren't tied to just one measure alone. What matters lies inside the FAST setup: it reshapes raw social

data into something models can actually use well. Layered timing cues get learned better than flat historical patterns, leading to sharper predictions every time. The gap between methods stays wide across tests, far beyond what random luck could explain.

D. Discussion

One clear takeaway stands out when thinking about public health forecasts. Not only does the data back up the main idea here, but it also shows how combined signs of mental state pulled from routine social posts may hint at rising self-harm patterns nationwide. What makes this different is not just spotting a link, instead turning that pattern into something more precise - a working forecast model shaped by digital behavior. Earlier studies had already noticed ties between what people express online and actual health shifts , yet now there's movement beyond mere connection toward repeatable estimation.

That ARIMA keeps performing poorly on both prediction jobs fits what we've heard before - models relying only on past numbers struggle when human behavior and shifting social factors shape outcomes, things official data often miss. What stands out is how a basic Decision Tree managed to beat fancier systems in MAE, just by using the time-organized results from the FAST process. This hints that getting the right setup and meaningful inputs can outweigh having a flashy algorithm.

Most models showed death predictions had worse accuracy compared to injury estimates. That makes sense - deaths happen less often, numbers jump around each month, plus patterns get lost in randomness. Looking at local population details or using tighter time intervals might improve forecasts down the line. The FAST system still proves useful nationwide, especially when official data comes late or misses pieces.

V. CONCLUSION

One reason this research began? A missing piece in how countries track self-harm: forecasts usually come too late, rely only on slow medical reports. To fix that, a new method called FAST came into play - tested, checked, shown to work. Instead of waiting months for hospital data, it pulls patterns from public conversations online about mental health. These digital traces, when grouped and studied, actually sharpen predictions. Not just vague guesses - they help point closer to real numbers of future self-harm injuries and deaths across whole nations.

From a test run in Thailand came some clear takeaways. Notably, models using social media cues beat standard methods every time forecast goals were set, especially XGBoost, which cut prediction mistakes nearly in half - about 43% less off track for deaths, just as close for injuries. What stands out is how Decision Trees, offered here as a new angle, missed the least in exact number guesses, proving simpler systems can hit hard when fed smart timelines. Oddly enough,

picking up signals without needing specific languages means this setup slips easily into different countries, barely needing tweaks at all.

This study adds to what we already know: how people act on social media can help predict public health trends in the real world. Instead of just noticing patterns, it pushes things forward by turning those links into working forecast tools. The FAST system gives officials a way to track health shifts quicker and easier than older methods. That matters most where medical records are spotty, uneven, or take too long to arrive - especially across parts of the Global South.

Even though outcomes look good, some limits need attention. This work focuses only on Thailand - although the method aims to be widely useful, testing elsewhere hasn't happened yet. Instead of pulling data from many platforms, researchers looked just at Twitter posts, a source that often misses voices of elderly people or those offline. Because government health records come out once a month, the analysis followed that rhythm, possibly overlooking sharp rises in self-harm cases between updates.

One path forward involves testing the method beyond just one country, trying it out across different places while adding platforms like Facebook or Reddit into the mix. Moving away from monthly summaries toward weekly check-ins might catch sudden changes in public mood more effectively. Instead of sticking only to what is already measured, bringing in local stats, job numbers, or how news feels emotionally could

sharpen predictions, especially when guessing something as difficult as deaths. What comes next might also include swapping traditional math models for newer ones built on deep learning, say those using memory-heavy designs or attention-focused patterns that track time differently. Trying these ideas wouldn't break from the original approach but grow naturally from it.

ACKNOWLEDGMENT

Thanks go straight to the teachers and mentors at the Computer Science and Engineering Department, CMR Technical Campus - each suggestion they offered quietly steered how this project unfolded. Without the support from department personnel behind the scenes, access to working machines and stable systems wouldn't have been guaranteed. Credit also lands with those who build and update free tools like Python, Scikit-learn, XGBoost, CatBoost, and Statsmodels; their code became the base layer here. Real progress only followed once actual numbers came into view - the self-harm records shared openly by Thailand's Ministry of Public Health gave everything a grounding point.

REFERENCES

[1] 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.

[2] 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.

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

[4] Y. S. Chang and J. Lee, "Is forecasting future suicide rate possible? Application of experience curve," *Eng. Manag. Res.*, vol. 1, no. 1, pp. 10, 2012.

[5] B. E. Belsher, D. J. Smolenski, L. D. Pruitt, N. E. Bush, E. H. Beech and D. E. Workman et al., "Prediction models for suicide attempts and deaths: A systematic review and simulation," *JAMA Psychiatry*, vol. 76, no. 6, pp. 642–651, 2019.

[6] J. M. Barros, R. Melia, K. Francis, J. Bogue, M. O'Sullivan and K. Young et al., "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, pp. 3201, Sep. 2019.

[7] 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.

[8] 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.

[9] A. E. Aiello, A. Renson and P. Zivich, "Social media- and internet-based disease surveillance for public health," *Annu. Rev. Public Health*, vol. 41, pp. 101, Apr. 2020.

[10] F. Feng, Y. Yang, D. Cer, N. Arivazhagan and W. Wang, "Language-agnostic BERT sentence embedding," *arXiv:2007.01852*, 2020.

[11] 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*, pp. 4040–4054, Jul. 2020.

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

[13] 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.

[14] 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.

[15] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang and H. Cho, "XGBoost: Extreme gradient boosting," vol. 1, pp. 1–4, 2015.

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

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