PUEY UNGPHAKORN INSTITUTE
FOR ECONOMIC RESEARCH

Using Large-Scale Social Media Data for Population-Level Mental Health Monitoring and Public Sentiment Assessment: A Case Study of Thailand

by

Suppawong Tuarob, Thanapon Noraset, and Tanisa Tawichsri

January 2022

Discussion Paper

No. 169

Using Large-Scale Social Media Data for Population-Level Mental Health Monitoring and Public Sentiment Assessment: a Case Study of Thailand

Final Report Submitted to Puey Ungphakorn Institute for Economic Research

October 31, 2021

by

Suppawong Tuarob, Thanapon Noraset, and Tanisa Tawichsri

Contents

Abstract

Mental health problems are among major public health concerns during the COVID-19 pandemic, given heightened uncertainties and drastic changes in lifestyles. However, mental health problem prevention and monitoring could be greatly improved given advancements in deep-learning techniques and readily available social media messages. This research uses deep learning algorithms to extract emotion, mood, and psychological cues from social media messages and then aggregates these signals to track population-level mental health. To verify the accuracy of our proposed approaches, we compared our findings to the actual number of patients treated for depression, attempted suicides, and self-harm cases reported by Thailand's Department of Mental Health. We discovered a strong correlation between the predicted mental signals and actual depression, suicide, and self-harm (injured) cases. Finally, we also create a database and user-friendly interface to facilitate researchers and policymakers to explore our extracted mental signals for further applications such as policy sentiment assessment.

Index Terms

Mental Health, Natural Language Processing, Deep Learning, Social Networks

I. Introduction

Given high uncertainties, job losses due to economic downturns, and lifestyle changes due to various measures such as quarantine protocols and lockdown during the COVID-19 pandemic, mental health problems become more prevalent across the world [1], and could have long-term impacts if left unaddressed. Mental health problems could have negative impacts both directly on population well-being and indirectly on economics due to productivity losses [2], [3]. Thus, public health authorities have established various indicators to monitor population mental health, such as rates of suicide, depression, alcoholism, and drug abuse. In Thailand, data for these indicators are collected through national healthcare systems such as the Thai Health Data Center[1]. While the indicators are helpful indicators for mental health monitoring, they are retrospective, and their applications may be limited. This study proposes to harness large-scale social media data and deep-learning techniques to create novel population mental health indicators that are proactive and powerful to timely gauge public sentiments and moods in response to policies.

The ability to monitor population-wide mental health and evaluate policy impacts on public sentiments could enable policymakers to assess public sentiments in response to the policy promptly and perhaps even adjust policy appropriately. Furthermore, policymakers can also provide mental health supports to alleviate the situations in a timely and proactive manner. For example, upon learning that a majority of people show a strong sign of stress in response to the announcement of the lockdown policy, the government could implement alleviation protocols, such as targeted psychological consultation programs that would cope with the upcoming possible mental issues in certain high-risk communities.

Due to the under-utilization of mental health services in Thailand, proactive strategies in providing targeted mental health supports and other ancillary programs that alleviate mental distress are needed. The proposed rapid index can help related agencies such as the Department of Mental Health coordinate with social workers, primary health care providers, or other government agencies already in contact with high-risk groups to proactively increase accessibility to mental health support services. Monitoring a period of heightened activity on the topic of mental health issues can also allow the public agencies to disseminate health information during the period of heightened interests for maximal impact [4]. Additionally, the successful development of algorithms with an accurate characterization of the high-risk groups and successful identification of early signs of mental health can lead to precise targeting technology using online platforms to increase mental health service usages in the future.

Online communities generate more than 2.5 quintillions (10^{18}) bytes of data world-wide each day [5], [6]. In 2017, 39.63 million Facebook users in Thailand were reported active[2]. This number is expected to grow to 45 million users by 2026. A large portion of this data is generated through social media services such as Twitter, Facebook, and Google Plus that process anywhere between 12 terabytes (10^{12}) to 20 petabytes

[1] https://hdcservice.moph.go.th

[2] https://www.statista.com/statistics/490467/number-of-thailand-facebook-users/

(10^{15}) of data each day [7]. These social media platforms allow their users to exchange information in a dynamic and seamless manner almost anywhere and anytime. Social media not only acts as a means of communication, but knowledge extracted from such large-scale social media data has also proven valuable in a wide variety of applications. For example, real-time analysis of Twitter and Facebook data has been used to monitor public healthcare [8], [9], simulate infection dynamics of infectious diseases [10], implement earthquake warning detection systems [11], predict the financial market movement [12], and identify notable product features [13]–[18].

One prominent use of social media is to serve as a timely and cheap alternative means to reflect real-world phenomena [10], [19]. In the public health domain, many studies have investigated the use of social media to monitor real-world population-wide and individual health, including epidemics [9], drug abuse [20], and identification of medical and emergency needs during the recovery from natural disasters (such as the Haiti Earthquake) [21]. The advancement of artificial intelligence and natural language processing technologies have enabled interpretation of the users' mental states while composing social media messages and have been widely adopted in many sentiments and mental health monitoring applications [22]. For example, automatically crawled Twitter data has been used to identify public mood towards specific events [23], suicidal thoughts [4], depression [24], [25], and other mental health signals [26].

Social media has also been investigated in economics and political science for a viable platform to study policy impact on population sentiment and mental health. For example, Twitter data was used to study the impact on public sentiment in response to the anti-immigrant laws [27], the Brexit [28], medical intervention policies [29], crimes [30], elections [31], [32], demonetization policies [33], trade policies [34], transportation policies [35], immigration policies [36], etc. Furthermore, it was shown that it is possible to infer public mental health from social media and sentiment signals [37]. Most of these studies investigate how the population responds to the policies or political events by analyzing their aggregate sentiment extracted from the tweets. Previous studies also showed that the analysis from public sentiment could be used to craft new or adjust existing policies [38], [39].

However, these methods were developed specifically for social media messages composed in English and for countries whose English is the main communication language, primarily due to the mature natural language processing algorithms and tools for such a high-resource language. To apply similar ideas in Thailand's settings, the research directions have faced tremendous limitations, namely:

- **Limitations of NLP Techniques**: Sentiment extraction requires advanced natural language processing (NLP) techniques to perform the following tasks: Part-of-Speech Tagging, Named Entity Recognition, Entity Resolution, Action Disambiguation, Semantic Inferencing, and Sentiment Feature Extraction. These tasks have been well studied for high-resource languages such as English; however, their development for low-resource languages such as Thai is still in its infancy.
- **Limitations of Data**: This research requires data in the form of online social media in Thailand, which has not been systematically collected and made available for research purposes. Automatically collecting such proprietary data could be challenging, mainly due to integrating data from heterogeneous platforms and handling incomplete/partial data.

Mental health and sentiment indexes constructed from organic big social network data can be compiled quickly and could offer more timely and cost-effective indicators than surveyed or administrative data. The key application of these indexes will be Nowcasting [40] of public mental health itself. If the predicted mental health problems are serious, related agencies can prepare to offer interventions such as hotlines and support groups, or increase mental health service capacity accordingly. Other applications include predicting stock indexes [41], job losses, or macroeconomics indicators like GDP [42]. These extracted social signals could also supplement official indexes, such as consumer sentiment or consumer confidence indexes. Specifically, the social media signals may have co-movements with these survey-based indexes but provide incremental information as the underlying population, timing, and data-generating processes are different. Furthermore, the ability to filter data by keywords and query processing will allow easy customization of indexes later on for specific issues.

This report investigates cross-lingual methods to estimate population-level mental health using predicted signals from social media messages. While the topic of mental health estimation using social media data has been previously studied, we examine a novel yet practical setting that is widely applicable – building such systems for poor-resource languages. In addition, considering that most cross-lingual work focuses on improving the performance of models, this article presents a unique downstream evaluation of such methods. We show that a language-agnostic deep learning model can be applied to create accurate population-level mental monitoring tools. The following is the main contributions of this report:

1) We present cross-lingual methods that leverage data sets in rich-resource language to build accurate text classification models for a poor-resource language.
2) We conduct comprehensive experiments to evaluate the cross-lingual methods on three mental-health-related tasks: sentiment classification, emotion classification, and suicidal tendency prediction.
3) We demonstrate that aggregated mental health signals from social media message classifications have high correlations with large-scaled ground-truth mental health data surveyed by the Ministry of Public Health.

The rest of this report is organized as follows. We first discuss related work on population-level mental health estimation and cross-lingual classification methods in Section II. In Section III, we present the methodology of the research, including classification algorithms and population aggregation of metal signals, followed by data sets, results, and discussion in Section IV. Finally, we conclude our work in Section VI.

II. Related Work

While monitoring population-level mental health using data from social media in Thailand has not been studied prior to this work, it and relevant methods have been studied in other settings. This section discusses advancements in social media mining, existing work on mental health monitoring, and cross-lingual approaches that help overcome our unique challenge.

A. Social Media Mining

Social media is now widely regarded as a key source of information for people to upload their generated content such as text, videos, photographs, and reviews. Therefore, social media data provides a wealth of information on people's thoughts, feelings, moods, and experiences throughout time, making it an ideal data source for mental health monitoring. We first give an overview of studies using traditional machine learning, then recent works using deep learning techniques are presented. Resnik et al. [43] exploited Latent Dirichlet allocation (LDA) to uncover underlying structure in collections of Twitter documents and studied meaningful linguistic signals for depression detection. Benton et al. [44] proposed a multi-task learning framework to predict various stages of mental health conditions. Burnap et al. [45] proposed to use Rotation Forest and a maximum probability voting to identify suicidal Tweets. Mowery et al. [46] classified depressed tweets by extracting lexical features and selecting the most discriminate features for the prediction. Chen et al. [47] studied the capability of using Support Vector Machine (SVM) and Random Forest (RF) to classify four types of mental disorders. Weerasinghe et al. [48] utilized LDA, SVM, a Bag-of-Words (BOW), and word clusters techniques to extract the feature from Tweets to achieve the depression classification task. Coppersmith et al. [49] leveraged the predictive power of Glove, bidirectional Long Short-Term Memory (LSTM), and attention network to obtain the most distinctive terms for suicide ideation prediction. Verma et al. [50] proposed to use a hybrid model of Convolutional Neural Network (CNN) and LSTM for detecting depressed tweets. Cong et al. [51] proposed to use attention bidirectional LSTM to extract informative terms, then XGBoost classifier is used to predict the class of depression symptoms. Tadesse et al. [52] exploited the word2vec technique with the LSTM and CNN model hybrid to predict the probability of suicidal messages on the Reddit platform.

B. Monitoring Mental Health via Social Media

Using social media to monitor mental health of users has been extensively studied. The majority of work focuses on identifying individual users or messages related to mental health such as depression, stress, or suicide [53]–[56]. Only limited research has demonstrated the ability of social media data to perform large-scale monitoring of important public mental health metrics such as population-level depression and suicide. A few that do often validate the monitoring systems by identifying abnormalities in a time series of the mental health signal and associating them with key events. For example, the system proposed by McClellan et al. can detect unexpected increases in the number of Tweets related to depression and suicide [4]. They also found that important events can explain such unexpected increases. Recent work by Zhou et al. studies how depression signals extracted from Twitter respond to key events or government policies related to COVID-19 [57]. On the other hand, research in other areas that analyses the population-level trends usually cross-validate with ground-truth statistics to assess the reliability of the predicted signals. For example, Google Search Index for relevant keywords has been found to correlate with influenza cases [58] or unemployment rate [59].

Since we would like to extract mental signals from social media messages, the problem could be framed as a short text classification task. Various text classification techniques have been studied to identify messages related to mental health. For example, McClellan et al. uses hand-crafted lists of hashtags and keywords [4]. As discussed in Section II-A, the majority of the previously proposed methods employ more data-driven approaches using machine learning models to improve the accuracy of individual prediction. While machine learning approaches avoid manually hand-crafted keywords, they require manually labeled corpus to build prediction models. This is the main challenge to overcome when deploying mental health monitoring systems in other languages where resources are not abundant, like in English.

C. Cross-lingual Text Classification

One of the most costly processes in building a classification model is obtaining a labeled dataset. While researchers have been regularly publishing relevant datasets, most of them are in English, and rarely in other languages. For example, in one of the largest collections of NLP datasets, *Datasets* library, there are 709 datasets in English, but only 45 datasets in Thai [60]. The NLP community understands this limitation and introduces text classification algorithms that exploit the abundance of resources in English to work with other low-resource languages. Hence, Cross-Lingual Text Classification (CLTC) aims to create a text classification model by using labeled data from a source language, and a little to no labeled data from target languages [61].

A simple method to leverage English resources for a low-language resource is to translate texts using a machine translation system. Many previous methods have demonstrated different approaches to using a machine translation system in the CLTC setting. Mihalcea et al. [62] and Banea et al. [63] showed that an annotated Romanian resource could be automatically created from an annotated English resource using a machine translation system. Instead of translating from the source language to the target language, Wan proposed that a Chinese text can be translated and input into an English sentiment analysis model [64]. Additionally, Salameh et al. confirmed this direction in the context of Arabic social media texts [65], [66]. They showed that a hand-crafted English sentiment analysis model using bag-of-words and part-of-speech representation is more robust than human annotators when predicting sentiments of the translated texts. With the advancement in text classification and machine translation research, we believe it is worth revisiting the machine translation approach for our mental health study.

At the foundation of modern text classification systems lies a pre-trained neural network such as word embeddings, recurrent language models, and recent transformer language models [67]–[72]. The pre-trained neural networks usually encode an input sentence into a vector representation so that a text classification can be effectively learned from a labeled corpus. This includes multi-lingual sentence encoders that work with many languages [71], but found to be under-performed in the CLTC setting [73], [74]. Perhaps, unsurprisingly, cross-lingual representation methods have been recently gained traction to improve upon

the multi-lingual encoders. Conneau et al. introduced a cross-lingual textual entailment corpus (a kind of text classification task) and later showed that a cross-lingual sentence encoder outperforms the machine translation approaches [75], [76]. Recently, Feng et al. introduced a cross-lingual training method that builds an effective language-agnostic language model (LaBSE) [77]. Renjit and Idicula trained LaBSE-based classification models using a limited resource of tweets in Dravidian Languages [78]. Gencoglu trained LaBSE-based classification models using a multi-lingual corpus to conduct a large-scale study on COVID-19 discourse [79].

Distinct from previous work, in this work, we focus on English as a source language and Thai as a target language. Although some previous works study social media texts in other languages, we create and evaluate classification for Thai tweets by having no labeled corpus for training in the target language. We consider various mental health signals extracted from the social media text as essential prediction evidence and experiment with both traditional and modern NLP techniques for textual modeling. Besides, the originality of our study lies in the pioneer in mining the social media messages for monitoring population mental health in Thailand. Finally, we quantitatively validate our system by examining the correlation of the predicted mental health signals with the ground-truth data published by the Thai government.

III. Methodology

The main goal of this study is to investigate whether mental health signals from social media can be used to estimate population-level mental health in Thailand, given that relevant annotated datasets are not available. The overall approach is to first extract mental health signals from individual messages and then aggregate the signals to represent the population-level mental health. Finally, we analyze the correlation between the aggregate mental health signals and the ground-truth mental health data. We formulate the mental signal prediction as a cross-lingual text classification because most languages, including Thai, the subject language of this article, do not have sufficient resources to build an accurate predictive model in the traditional setting. When estimating population-level mental health, we use all messages collected from social media to aggregate their predicted mental health signals over the time dimension. Since there are many potential mental health signals, we only choose mental health signal related to depression and suicide attempts. The following subsections discuss each component of our approach in detail.

A. Mental Signal Tasks

Our underlying task is mental signal extraction, where a text document is labeled into a category related to mental states. We could directly model key mental health signals such as depression and suicide attempts for individual documents. Unfortunately, we have limited data sets that we can sufficiently build text classification models. Given limited availability of the datasets, however, we study three related mental signal tasks: *Emotion*, *Sentiment*, and *Suicidal tendency*. We use the mental signal tasks in the three data sets as a proxy for the important mental health signals. We expect that the aggregate statistics of individual social messages classified as negative emotions, negative sentiments, or high suicidal tendency might correlate with country-level depression and suicide attempts. The three mental signal datasets are in English and their relevant statistics are shown in Table II.

Emotion: The first dataset, *Emotion* comes from the GoEmotions project [80]. The GoEmotions dataset consists of roughly 54,000 English messages from Reddit. Each message is manually categorized into 27 labels of fine-grain emotions. The labels are too detailed for our application, and it would be difficult to translate into other cultures. Hence, we convert the fine-grain emotion labels to Ekman emotions [81] using the mapping provided with GoEmotions data. The Ekman emotions consist of anger, disgust, fear, joy, sadness, surprise, and neutral.

Sentiment: In addition to the emotion signals, we study sentiment signals. The GoEmotions also provide a mapping from the fine-grain emotion labels to sentiment labels. Our second data set, *Sentiment*, is the sentiment-mapped GoEmotions data set. The sentiment signals include positive, negative, ambiguous, and neutral.

Table I
Acronyms used in this paper.

Acronym	Original Term	Type
ME-Ang	Anger	Mental Signal
ME-Dis	Disgust	Mental Signal
ME-Fea	Fear	Mental Signal
ME-Joy	Joy	Mental Signal
ME-Sad	Sadness	Mental Signal
ME-Sur	Surprise	Mental Signal
ME-Neu	Neutral (Emotion)	Mental Signal
MS-Pos	Positive	Mental Signal
MS-Neg	Negative	Mental Signal
MS-Amb	Ambiguous	Mental Signal
MS-Neu	Neutral (Sentiment)	Mental Signal
M-ST	Suicidal	Mental Signal
M-NST	Non-suicidal	Mental Signal
GP-Depress	# Depression Patients	Ground-Truth Administered Data
GH-Death	# Successful Suicidal Attempts	Ground-Truth Administered Data
GH-Visit	# Potential Suiciders	Ground-Truth Administered Data
GH-Injure	# Non-Successful Suicidal Attempts	Ground-Truth Administered Data
MNB	Multinomial NaiveBayes	Classification Model
SVM	Support Vector Machine	Classification Model
BiLSTM	Bi-Directional Long Short-Term Memory (LSTM)	Deep Learning Classification Model
BERT	Bidirectional Encoder Representations from Transformers	Deep Learning Classification Model
RoBERTa	Robustly optimized BERT approach	Deep Learning Classification Model
LA-BERT	Language-Agnostic BERT	Deep Learning Classification Model

Suicidal Tendency: Our third data set is collected from r/SuicideWatch subreddit inspired by Shing et al. [82] and later the CLPsych 2019 Shared Task [80]. Specifically, we collect posts from r/SuicideWatch subreddit. These posts are self-reports of Reddit users who experience suicidal thoughts and self-harm. Thus, we consider these as messages of high suicidal tendency. We, then, select positive emotion messages from GoEmotions as negative samples (non-suicidal) to complete *Suicidal Tendency* data set. The suicidal tendency is a binary signal (suicidal or non-suicidal).

B. Cross-lingual Text Classification

All three datasets we have are in English, but our target social media messages are in Thai. This is a setting of the cross-lingual text classification (CLTC). The CLTC aims to leverage a model trained using labeled data from one language and used to classify documents of a new language without manually labeling data in the new languages. More formally, we have a set of labeled documents in a source language $D^S = \{(x_1^S, y_1), (x_2^S, y_2), ..., (x_N^S, y_N)\}$, where x^S is a document in the source language and $y \in \{0, 1\}$ is the label signal. In addition, we have a set of unlabeled documents in a target language $D^T = \{(x_1^T, x_2^T, ..., x_M^T)\}$. We assume that D^S and D^T are not parallel corpora and can come from different sources, i.e., different social media platforms. Our goal is to use labeled documents in the source language D^S to train a classifier that predicts labels (y) for an unseen set of target-language documents D^T.

In this work, we investigate two CLTC methods: machine translation and cross-lingual representation. In both CLTC methods, an input document is first encoded to extract a feature vector v by a document

Table II
Training data sets (composed in English) for emotion [80], sentiment [80], and suicidal tendency [83] extraction models, as well as corresponding validation sampled data from real Tweets composed in Thai.

Dataset	Mental Signal	# Messages (English)	# Annotated Thai Tweets
Emotion	Anger (ME-Ang)	7,022	100
	Disgust (ME-Dis)	816	100
	Fear (ME-Fea)	883	100
	Joy (ME-Joy)	21,119	100
	Sadness (ME-Sad)	3,212	100
	Surprise (ME-Sur)	5,190	100
	Neutral (ME-Neu)	16,021	100
	Total* - Emotion	54,263	700
Sentiment	Positive (MS-Pos)	21,733	100
	Negative (MS-Neg)	11,319	100
	Ambiguous (MS-Amb)	5,190	100
	Neutral (MS-Neu)	16,021	100
	Total* - Sentiment	54,263	400
Suicidal Tendency	Suicidal (M-ST)	116,037	100
	Non-suicidal (M-NST)	33,052	100
	Total - Suicidal	149,089	200

encoder $\phi(\cdot)$, then a classification model $f(\cdot)$ predicts a probability of the labels from the feature vector:

$$v = \phi(x; \omega) \tag{1}$$

$$P(y|x) = f(v; \theta) \tag{2}$$

, where ω and θ are parameters for the document encoder and the classification model. In this work, we model each class of the mental signal independently to mitigate the problem of the imbalance class. Following the state-of-the-art approaches for text classification, we apply deep learning approaches for both $\phi(\cdot)$ and $f(\cdot)$ where the probability is given by the Sigmoid function. Then, we minimize the standard cross-entropy loss to learn both sets of parameters together for each class:

$$L(\omega, \theta; D^S) = \sum_{(x_i^S, y_i) \in D^S} y_i \log f(\phi(x_i^S; \omega); \theta) \tag{3}$$

Note that the training only uses D^S. Furthermore, the document encoder is usually pre-trained on other corpora prior to CLTC (i.e., the initial ω is given and fine-tuned using D^S).

Machine Translation Approach: In the machine translation approach, both a document encoder and a classification model are trained solely on the source language data. A document encoder is a pre-trained language model of the source language such as BERT [71] and RoBERTa [72]. To predict a label, a document x^T is first translated into a source language $\tilde{x}^S$, encoded into a document representation, and then fed into the classification model. Formally, given a machine translation system $\text{MT}^{T \to S}$, the predicted label $\hat{y}$ is computed as follow:

$$\tilde{x}^S = \text{MT}^{T \to S}(x^T) \tag{4}$$

$$\hat{y} = \mathbb{I}[P(y|\tilde{x}^S; \omega, \theta) > c] \tag{5}$$

, where $\mathbb{I}[\cdot]$ is an indicator function having the value of 1 if the condition is true, otherwise 0, and $c \in [0, 1]$ is a constant threshold (0.5 throughout this paper). $\text{MT}^{T \to S}$, however, might not accurately translate a document and confuses the pre-trained language model. To establish a baseline, we also investigate other text classification approaches in the experiments. These approaches include a bag-of-word encoder (TF-IDF) with a statistical learning algorithms (Multinomial Naive Bayes [84] and SVM [85]) and the bidirectional LSTM initialized with the FastText's word embedding [68], [86], [87].

Cross-lingual Representation Approach: In the cross-lingual representation approach, the document encoder is a pre-trained language model that can produce language-agnostic representation. In other words, a language-agnostic language model outputs similar representation for x^S and x^T that have similar meanings. Such pre-trained language-agnostic language models (such as LaBSE [77]) are trained using parallel corpora of multiple languages, including documents in the source language and the target language. The system can predict x^T directly without using a machine translation system:

$$\hat{y} = \mathbb{I}[P(y|x^T; \omega, \theta) > c] \tag{6}$$

Evaluation of CLTC Approaches: For both CLTC methods, we evaluate both held-out source-language documents and the target language documents. The source language evaluation uses the original data sets. The target language evaluation uses our manually-labeled data sets of Thai tweets. We collect the Thai Tweets from Twitter using keywords according to classes that we consider and use the three human raters to manually label these data sets and resolve the final labels by the majority votes. We only keep one hundred samples of each class. Table II shows numbers of total English documents for mental signal classes and the labeled Thai documents for testing.

C. Population-level Mental Health Estimation

Previous studies suggested that overall mental health signals extracted from social media messages can represent population mental health [56], [57]. This work estimates monthly population-level mental health indicators using the aggregate statistics of extracted signals from individual social media messages. Formally, given a collection of social media messages for m-th month, D_m^T, we represent the aggregate statistics of mental health signals $z_m^{(j)}$ by a proportion of individual messages classified as the label $y^{(j)}$:

$$z_m^{(j)} = \frac{1}{|D_m^T|} \sum_{x^T \in D_m^T} \hat{y}^{(j)} \tag{7}$$

We repeat this process for every month and every mental health signal to obtain monthly trends of all mental health signals (13 types in Table II). For each mental signal series, we compute a cross-correlation with the ground-truth mental health series to check the accuracy of the methods.

Cross Correlation with Time Shifts: To cross-validate the aggregate extracted mental health signals, we use the Pearson correlation coefficient between monthly ground-truth mental illness cases data and each aggregate mental health signal. High correlations would indicate that social media messages' aggregate mental health signals can accurately represent the population-level mental health. The relationship between people's behaviors on social media and observable mental health problem realization in real life is not yet well established (i.e., the dynamic between the two series, which precedes which, are still unknown). Thus, the correlation with varying time shifts (horizons), h, between the two series is computed. Formally, we compute the time-shifted correlation r_h^{jk} between a normalized monthly ground-truth mental health statistics ($g_m^{(k)}$) and a monthly aggregate mental health signal ($z_m^{(j)}$) as follow:

$$r_h^{jk} = \frac{\sum_m (z_{m+h}^{(j)} - \bar{z}^{(j)})(g_m^{(k)} - \bar{g}^{(k)})}{\sqrt{\sum_m (z_{m+h}^{(j)} - \bar{z}^{(j)})^2} \sqrt{\sum_m (g_m^{(k)} - \bar{g}^{(k)})^2}} \tag{8}$$

, where $\bar{z}^{(j)}$ and $\bar{g}^{(k)}$ are the averages of aggregate mental health signals and the ground-truth statistics respectively. When $h < 0$, the signal from social media is a post reflection, or *lags*, of ground-truth data. When $h > 0$, the signal from social media is a forecast, or *leads*, of ground-truth data.

Table III
data sets used to illustrate the applicability of the models, including sampled Tweets collected in Thailand and ground-truth administrative data (depression patients [GP] and suicidal cases [GH]), during the 2019 fiscal year (i.e., October 2019 - September 2020) in a total of 12 months.

Month (MM-YYYY)	# Case-Study Tweets	Ground-Truth Administered Data			
		GP-Depress	GH-Death	GH-Visit	GH-Injure
10-2019	226,219	10,066	434	180	1,121
11-2019	388,977	10,927	331	108	970
12-2019	102,375	86,136	393	288	1,021
01-2020	25,813	9,747	468	364	1,147
02-2020	16,544	8,418	429	249	1,023
03-2020	61,093	9,641	441	317	996
04-2020	71,228	3,646	376	310	843
05-2020	85,328	7,369	366	226	969
06-2020	46,505	8,102	400	256	869
07-2020	107,853	12,094	321	184	846
08-2020	78,357	11,324	332	135	840
09-2020	76,650	6,783	360	30	900
Total	1,286,942	184,253	4,651	2,647	11,545

Case-Study data sets: To evaluate the system's overall performance, we collect ground-truth mental illness cases data reported by Thailand's Department of Mental Health from October 2019 to September 2020 (one fiscal year in Thailand). We select two main health metrics that the Thai government collects and monitors: monthly numbers of patients that were diagnosed and treated for depression at hospitals and healthcare providers under the provision of the Department of Mental Health (GP-Depress)[3] and also monthly cases of patients that visited hospitals for suicide, suicide attempts, and self-harm (GH)[4]. The suicide cases are categorized into three subcategories: 1) suicidal patient visits (GH-Visit), 2) cases with unsuccessful suicide attempts (GH-Injure), and 3) successful suicide cases (GH-Death). The ground-truth mental health data are normalized by the number of Thai population in each year (2019 and 2020). We collect monthly representative samples of Twitter messages in the Thai language. Table III shows the monthly numbers of Tweets and the ground-truth administrative data. Note that we collect the Tweets from July 2019 to December 2020 (± 3 months) to accommodate the time-shifted analysis.

IV. Experiments and Results

We present our findings in three parts. First, we experiment on the English mental signal classification tasks to find the best algorithm for each mental signal task. Then, we experiment with the same tasks, but in the CLTC setting where the source language is English and the target language is Thai. Finally, we present the cross-correlations between the Thai ground-truth administrative data with our experiments' aggregate mental health signals.

A. Experiment Setting and Model Selection

To find the best algorithm for the mental signal classification tasks, we experiment with three main approaches: the traditional bag-of-words approach, the LSTM approach, and the language model fine-tuning approach. In the first approach, we encode a document using a TF-IDF vector in which we keep only 200 most frequent words. We built three TF-IDF document encoders for each dataset. Then, we use the TF-IDF vectors to train a Multinomial Naive Bayes model (MNB) [84] and a Support Vector Machine model (SVM) [85] for each dataset. In the LSTM approach, we trained a document encoder and a classification model together using bi-directional LSTM architecture [86], [87]. We initialized the word embeddings using an embedding set from FastText [68]. Finally, in the language model fine-tuning approach, we experimented with BERT [71], RoBERTa [72], and LaBSE [77]. We used AdamW optimizer [88] with the learning rate of

[3]https://www.thaidepression.com/www/report/main_report
[4]https://506s.dmh.go.th

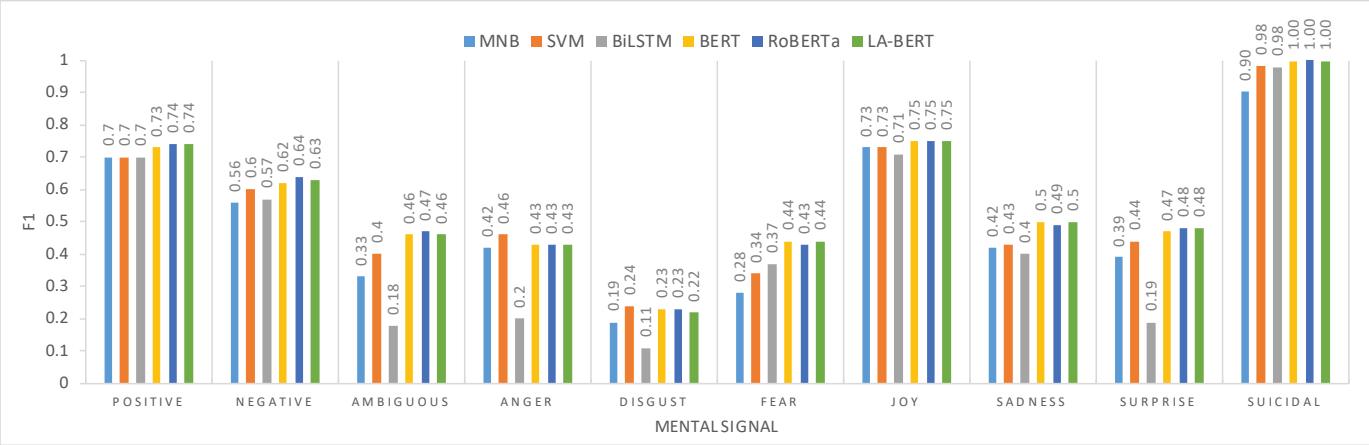


Figure 1. F1 comparison of different models on all the mental signal identification tasks.

2e-5, epsilon of 1e-6, weight decay of 0.1, and train for 4 epochs. To mitigate the class imbalance problem, we used SMOTE to synthesize more examples in minority classes [89].

The English mental signal datasets (in Table II) are split into 80:20 for training/testing. We report F1 scores for each mental signal class in Figure 1. We can see that the language modeling fine-tuning approach yields the best testing F1 score (BERT, RoBERTa, and LaBSE). These results align with the previous findings that find BERT-based models outperform the traditional bag-of-words and word-embedding approaches [80], [83]. Interestingly, the language-agnostic models (LaBSE) perform on par with the other English language models. It is also worth pointing out that all models perform well on the suicidal tendency data set but poorly on the fine-grain negative emotions, especially the class *Disgust* due to low recall (but high precision). With these results, we select BERT and LaBSE for further analyses.

B. Evaluation on Thai Social Media Messages

We evaluated the performance of the mental signal models on the manually labeled Thai tweets. We classified Thai messages using the CLTC approaches described in Section III. This experiment used the best-performing models from the previous experiment and compared the machine translation (BERT) and the cross-lingual representation (LaBSE) methods. We also include the traditional text classification approach using the machine translation method (SVM). The traditional text classification approach uses a bag-of-word representation that could be more robust to translation noise because it is agnostic to word orders. We reported and compared precision (P), recall (R), and F1 for all mental health signals to evaluate the model performance. The machine translation model from Thailand Artificial Intelligence Research Institute [90] is used to translate Thai Tweets to English messages. This machine translation model was reported to achieve the state-of-the-art performance on the IWSLT 2015 corpus (TED Talk transcripts) [91].

Tables IV, V, and VI show the performance of the three methods on the *Emotion*, *Sentiment*, and *Suicidal Tendency* datasets respectively. First, we can see that the traditional text classification approach (SVM) performs worse than the language model fine-tuning approach (BERT) in the machine translation setting. LaBSE has the highest F1 scores for overall performance, except for the *anger* and the *neutral* emotion signals. Since both LaBSE and BERT have similar F1 scores in the English evaluation, the result indicates that the cross-lingual representation approach is better than the machine translation approach in the CLTC setting.

While we can conclude that LaBSE is the best approach in our setting, we can observe differences in performance in terms of individual mental health signals. Since the mental health signals predicted by our best method yield a lower noise level in the aggregate statistics than the others, it is important to identify the mental health signals suitable for further analysis. In Table IV, the LaBSE model has high F1 scores for the *fear* and *sad* emotions and high precision for the *disgust* emotion. In Table V, there are

Table IV
Classification performance of the selected classifiers on the emotion detection task, validated with annotated Thai Tweets.

Model	Class	P	R	F1
SVM	ME-Ang	0.41	0.9	0.57
	ME-Dis	0.74	0.54	0.62
	ME-Fea	0.7	0.87	0.77
	ME-Joy	0.39	0.94	0.55
	ME-Sad	0.36	0.95	0.52
	ME-Sur	0.49	0.81	0.61
	ME-Neu	0.8	0.66	0.72
	Macro Avg	0.53	0.8	0.61
BERT	ME-Ang	0.62	0.82	0.71
	ME-Dis	1	0.11	0.2
	ME-Fea	0.76	0.81	0.79
	ME-Joy	0.44	0.98	0.61
	ME-Sad	0.57	0.82	0.67
	ME-Sur	0.54	0.77	0.64
	ME-Neu	0.8	0.66	0.72
	Macro Avg	0.68	0.71	0.62
LA-BERT (No Translator)	ME-Ang	0.57	0.42	0.48
	ME-Dis	1	0.53	0.69
	ME-Fea	0.94	0.87	0.9
	ME-Joy	0.5	0.96	0.66
	ME-Sad	0.67	0.83	0.74
	ME-Sur	0.57	0.82	0.67
	ME-Neu	0.68	0.73	0.7
	Macro Avg	0.7	0.74	0.69

Table V
Classification performance of the selected classifiers on the sentiment detection task, validated with annotated Thai Tweets.

Model	Class	P	R	F1
SVM	MS-Pos	0.44	0.47	0.46
	MS-Neg	0.61	0.7	0.65
	MS-Amb	0.08	0.21	0.11
	MS-Neu	0.65	0.83	0.72
	Macro Avg	0.44	0.55	0.49
BERT	MS-Pos	0.41	0.63	0.49
	MS-Neg	0.64	0.78	0.70
	MS-Amb	0.19	0.49	0.27
	MS-Neu	0.66	0.91	0.77
	Macro Avg	0.47	0.70	0.56
LA-BERT (No Translator)	MS-Pos	0.47	0.70	0.56
	MS-Neg	0.72	0.84	0.77
	MS-Amb	0.33	0.65	0.44
	MS-Neu	0.72	0.92	0.81
	Macro Avg	0.56	0.78	0.64

Table VI
Classification performance of the selected classifiers on the suicidal tendency extraction task, validated with annotated Thai Tweets.

Model	Class	P	R	F1
SVM	M-ST	0.84	0.67	0.74
	M-NST	0.73	0.87	0.79
	Macro Avg	0.78	0.77	0.77
BERT	M-ST	0.76	0.96	0.85
	M-NST	0.95	0.7	0.8
	Macro Avg	0.85	0.83	0.83
LA-BERT (No Translator)	M-ST	0.84	0.87	0.86
	M-NST	0.87	0.84	0.85
	Macro Avg	0.85	0.86	0.85

negative and *neutral* sentiments that the model performs well. Finally, in Table VI, the model achieves high performance for *suicidal tendency* prediction.

C. Correlation Analysis with Administrative Data

With the predicted mental health signals, we next evaluate how well our aggregate mental health signals can represent population mental health using correlations between the aggregate mental health signal statistics and ground-truth mental health data. Here, we have five mental health signals that have the F1 score higher than 0.7 but excluded the neutral emotion and sentiment (discussed in Section IV-B) and four indicators of population mental health (the number of actual mental health-related cases discussed in Section III-C). In addition to the five mental health signals, we use Google Search Index from Google Trend as another benchmark for comparison with our mental health signals to see how well our mental health signals perform compared with a state-of-the-art baseline. Google Search Index is chosen as a benchmark because it has been used in the literature as a proxy for epidemics [58] and depression indicators [59]. This study uses keywords: "depression" (in Thai) to cross-validate with GP-Drepress, and "suicide" (in Thai) to cross-validate with GH-Visit, GH-Injure, and GH-Death, respectively. To analyze the correlation, we compute Pearson correlation coefficients (r) with time shifts as defined by Equation 8 with $h \in \{-3, -2, -1, 0, 1, 2, 3\}$. Table VII presents only the maximum correlation coefficients for each pair of the administrative data and mental signal (the complete pair-wise coefficients are presented in Table IX).

Table VII
Pearson correlation coefficients (r) between administrative data and mental health signals from Thai tweets during October 2019 - September 2020; showing only the best time shifts (h).

Signal	GP-Depress		GH-Visit		GH-Injure		GH-Death	
	r	h	r	h	r	h	r	h
ME-Fea	0.555	1	0.577	3	0.823	2	0.869	3
ME-Sad	0.498	1	0.305	-3	0.810	2	0.746	3
ME-Dis	0.487	2	0.669	3	0.810	2	0.892	3
MS-Neg	0.657	3	-0.037	-3	-0.306	-2	-0.293	-2
M-ST	0.543	-2	0.769	-1	0.913	-3	0.856	-3
GT (baseline)	0.641	2	0.195	2	0.288	3	0.063	2

Our results show that the extracted mental signals align with population-level mental health statistics and performed better than the baseline Google Search Index as concurrent, lead, and lag indicators. Generally, many of the mental health signals from Thai tweets have positive correlations with the administrative data. Most mental health signals have leading trends ($h > 0$) to the ground-truth data and have a higher correlation than the baseline Google Search Index. Meanwhile, the baseline Google Search Index's concurrent and post reflection (h <0) sequences even negatively correlate with ground-truth data. Starting with the results for the number of depression patients (GP-Depress), negative sentiment signal (MS-Neg) has the highest correlation with the ground-truth depression data, i.e., the number of patients receiving treatment for depression each month. Surprisingly, MS-Neg has a low negative correlation with the suicidal data.

For the suicidal data (GH-Visit, GH-Injure, and GH-Death), the suicidal tendency signal (M-ST) has a strong positive correlation with all three types of administrative data. Distinct from the other signals, the M-ST signal lags behind the administrative data. On the other hand, the fear emotion signal (ME-Fea) also has a high positive correlation with the suicidal cases but leads the ground-truth data two-to-three months. Hence, the results indicate that the predicted mental signal aggregated from Thai tweets can represent the Thai population-level mental health metrics.

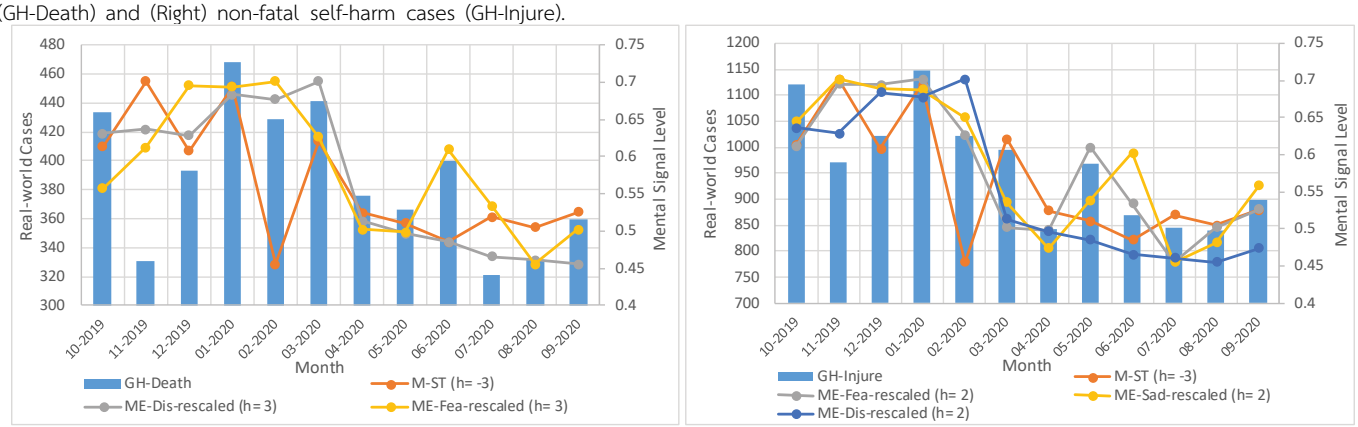
V. Discussion

Given that the dynamic and relationship between mental health patient data, social media, and Google Trend are not well documented, one of the main contributions of this study is to investigate the relationship

Table VIII
Monthly numbers of Tweets classified into different mental health signals during October 2019 - September 2020.

Month (MM-YYYY)	ME-Ang	ME-Dis	ME-Fea	ME-Joy	ME-Sad	ME-Sur	ME-Neu	MS-Pos	MS-Neg	MS-Amb	MS-Neu	M-ST	M-NST
10-2019	11,561	445	3,839	48,368	8,383	23,219	130,404	27,857	61,113	697	136,552	57,533	168,686
11-2019	22,226	744	6,782	70,635	14,353	44,653	229,584	39,505	106,786	1,697	240,989	96,560	292,417
12-2019	5,363	217	1,332	16,851	3,330	11,049	64,233	10,204	26,455	438	65,278	27,112	75,263
01-2020	569	13	147	3,990	496	2,309	18,289	1,755	3,754	66	20,238	15,841	9,972
02-2020	243	6	90	2,365	199	1,191	12,450	1,081	1,847	35	13,581	11,605	4,939
03-2020	2,517	16	735	7,092	1,197	7,046	42,490	3,381	13,739	181	43,792	37,115	23,978
04-2020	3,357	7	539	6,431	1,920	9,783	49,191	2,591	18,784	189	49,664	49,428	21,800
05-2020	1,245	5	249	5,026	846	5,369	72,588	2,020	17,647	99	65,562	38,874	46,454
06-2020	1,321	0	264	8,389	600	3,837	32,094	6,793	7,589	97	32,026	28,863	17,642
07-2020	3,334	18	764	29,584	2,354	9,375	62,424	26,250	28,711	99	52,793	56,556	51,297
08-2020	1,854	11	410	25,468	1,593	6,319	42,702	22,466	16,899	99	38,893	39,981	38,376
09-2020	1,832	15	411	23,990	1,501	6,093	42,808	20,952	15,262	80	40,356	37,225	39,425

Figure 2. Illustration of some notable mental health signals that are highly correlated with the real-world (Left) successful suicidal attempts (GH-Death) and (Right) non-fatal self-harm cases (GH-Injure).



between those three data sources. The three main data sources in the correlation analysis exercises are the following. 1) The ground truth data: the real-world data reflecting the number of patients with depression, suicidal tendency, and suicide attempts (both successful and unsuccessful) that received healthcare treatment. 2) Mental signals extracted by our deep-learning tools from sample Tweets in Thai languages. And 3) Google Trend: the Google Search Indexes on selected terms ("depression" and "suicide") as a baseline proxy of related mental health problems. This section will first discuss the underlying population of each data source and then the dynamic relationship between these data sources observed from the correlation analysis with time-shift sequences.

The underlying populations of the three data sources are different. The ground truth data shows the proportion of the Thai population receiving mental health treatment on depression and suicide-related treatment. Meanwhile, social media data and Google Trend generate mental health proxy of users on respective platforms: Twitter and Google. In January 2021, there were about 49 million internet users in Thailand, equivalent to 70% Internet penetration rate, and most of the internet users use Google[5]. According to the latest data reported by Statista in 2020, more than 90% of internet users now use social media. Social media usages are the highest in the Generation Y age group (97.3%) and lowest among the baby boomers (89.4%).

Aside from different population coverage, the following discusses the underlying population of each data source in more detail. First, ground truth data of depression patients only reflects those who have access to healthcare, decide to get professional help about their mental health problems, and were diagnosed with depression. Thus, the patients appearing on the administrative data are likely to have more severe

[5]https://www.wikigender.org/

mental health problems and might be better off with access to healthcare. The suicide-related patients also likely have more severe mental health problems than from other sources, given that they are diagnosed with suicidal tendencies or already attempted suicides.

In comparison, people who have depression or suicidal tendencies on social media or searches about depression or suicide on Google could have ranging severity in mental distresses. They might not have mental health problems that are serious at the clinical level yet. However, they are interested in the selected keywords ("depression" and "suicide") and might simply just self-diagnose themselves [92]. Moreover, they may not have professional treatment for various reasons, such as not having access to healthcare, being afraid to get treated professionally, or not believing the treatment could help them. Comparing Google and social media users, the underlying population for both Google and social networks must have access to the Internet and possess sufficient digital literacy to use those platforms. For this study, social media data are from Twitter, which may have younger users than Google.

Besides different underlying populations, mental health problems manifested and observed in each data source may reflect different stages of mental illnesses in the real world. Therefore, one data source may be a concurrent proxy, post reflection (lag indicator), or forecasting variable (lead indicator) of others. For instance, the ground truth data of depression or suicide-related patients only reflects those with relatively severe mental health problems seeking professional health and having access to such healthcare. However, the reflection of mental health problems in social media messages could manifest even before people realize that they have depression or suicidal tendencies, having depression or suicidal tendencies shown in the rhetoric or languages detected in their social messages. However, from social media alone, it is unknown that people with depression and suicidal tendency signals are already aware of their mental problem and are receiving healthcare treatments or not. Meanwhile, most people searching about depression or suicides most likely are already aware of their own potential mental illness and might be self-diagnosing. Thus, the proxy from Google Trend could reflect those with awareness of mental health issues but may or may not be seeking and receiving professional healthcare treatment.

Compared to mental signals from social media and the google trend, mental health problems manifested in the patient data are likely to be more severe and at later stages. Mental signals from social media data could be from any stages of mental illnesses: 1) even before the subjects are aware of their mental health problems, 2) already aware but not yet seeking or receiving treatment, or 3) after they already receive professional treatment. However, mental health proxy picked up by Google Trend will likely be after they are already aware of their own mental health issues but they may or may not be receiving professional treatment. However, there could also be some noises in Google Trend when the indexes might surge if there is trending news related to the keywords.

Our correlation analysis found relationships between three data sources to be in line with the expected dynamic relationship. Given the underlying relationship of each data source to stages in mental illnesses discussed above, one can expect Google Search Index to be a lead or concurrent indicator to the patient data if the underlying population is self-diagnosing. However, Google Trend could also be a lagged indicator if trending news related to the keywords triggers the surge in keyword searching. It could be both lead or concurrent indicators for social media data but less likely a lagged indicator. From Table VII, Google Search Index is shown to be a lead indicator to the ground-truth patient data, with a positive correlation in the forecasting time shift (h>0) sequences, while the concurrent and lags have a negative correlation with ground truth data. However, mental signals extracted from social media data are shown to have a stronger positive correlation to ground truth data than Google Trend.

VI. Conclusions and Future Directions

Using social network data, this research demonstrates how the detection of population-level mental states may be utilized to explain the aggregated level of mental health in the population in Thailand. Thailand's social media messages were used to train deep learning-based algorithms to identify mental health symptoms. Cross-validating the extracted signals with real-world depression, suicide, and self-harm

cases during the fiscal year of 2019, we found a strong correlation between certain signals such as suicidal tendency, fear, sadness, and disgust. We believe the findings can also be utilized to offer more timely input on government initiatives. Future research will extract and analyze the main events from social media communications that signal a dramatic shift in the number of mental illness cases.

We also found that the extracted mental signals from social media are lead and concurrent indicators to depression and suicides and are stronger indicators than Google Trend. The result is promising and encouraging that our algorithms could have real-world applications in the early detection of mental health problems from social media data (or, more generally, from languages people use). However, using only aggregate data, we cannot determine whether the misalignment between two data sources is from noise or classification models' errors. Future studies with individual-level data with matched social media data and mental health history should be explored to further improve the accuracy and verification of extracted mental signals.

Appendix
Complete Correlation Analysis

Table IX
Pearson correlation efficient between administrative data and top-3 mental health signals during October 2019 - September 2020, varying horizons from $h = -3$ (3-month reflection) to $h = 3$ (3-month forecasting).

Reflectability	Horizon	Signal	GP-Depress	Signal	GH-Death	Signal	GH-Visit	Signal	GH-Injure
Post Reflection	-3	MS-Neg	0.360	M-ST	0.856	M-ST	0.648	M-ST	0.913
	-3	ME-Ang	0.287	MS-Neg	-0.478	MS-Neg	-0.037	MS-Amb	-0.564
	-3	ME-Fea	0.274	ME-Ang	-0.660	ME-Ang	-0.291	MS-Neg	-0.568
	-3	GT (DP)	-0.101	GT (SC)	-0.195	GT (SC)	-0.225	GT (SC)	-0.268
	-2	M-ST	0.543	M-ST	0.601	M-ST	0.595	M-ST	0.575
	-2	ME-Fea	-0.321	MS-Neg	-0.293	MS-Neg	-0.302	MS-Amb	-0.197
	-2	ME-Sad	-0.444	MS-Amb	-0.341	MS-Amb	-0.371	MS-Neg	-0.306
	-2	GT (DP)	-0.214	GT (SC)	-0.469	GT (SC)	-0.437	GT (SC)	-0.599
	-1	M-ST	0.502	M-ST	0.588	M-ST	0.769	M-ST	0.420
	-1	MS-Amb	-0.415	MS-Amb	-0.015	MS-Amb	-0.323	MS-Amb	0.289
	-1	ME-Dis	-0.420	ME-Ang	-0.224	MS-Neg	-0.504	ME-Ang	-0.069
	-1	GT (DP)	-0.282	GT (SC)	-0.443	GT (SC)	-0.155	GT (SC)	-0.412
Current Prediction	0	M-ST	0.166	MS-Amb	0.290	M-ST	0.438	ME-Dis	0.537
	0	MS-Amb	0.082	M-ST	0.266	MS-Amb	0.287	MS-Amb	0.475
	0	ME-Ang	-0.119	ME-Dis	0.085	ME-Ang	0.010	ME-Fea	0.385
	0	GT (DP)	-0.282	GT (SC)	-0.155	GT (SC)	-0.247	GT (SC)	0.024
Forecasting	1	ME-Fea	0.555	MS-Amb	0.405	MS-Amb	0.636	ME-Dis	0.726
	1	MS-Amb	0.526	ME-Dis	0.310	ME-Dis	0.302	MS-Amb	0.698
	1	ME-Sad	0.498	ME-Fea	-0.008	ME-Fea	0.251	ME-Sad	0.552
	1	GT (DP)	0.518	GT (SC)	-0.190	GT (SC)	-0.242	GT (SC)	0.212
	2	ME-Fea	0.512	MS-Amb	0.770	MS-Amb	0.771	ME-Fea	0.823
	2	ME-Sad	0.488	ME-Sad	0.727	ME-Dis	0.560	ME-Sad	0.810
	2	ME-Dis	0.487	ME-Dis	0.694	ME-Fea	0.528	ME-Dis	0.810
	2	GT (DP)	0.641	GT (SC)	0.063	GT (SC)	0.195	GT (SC)	0.018
	3	MS-Neg	0.657	ME-Dis	0.892	ME-Dis	0.669	ME-Fea	0.665
	3	ME-Ang	0.345	ME-Fea	0.869	MS-Amb	0.630	ME-Sad	0.654
	3	ME-Sad	0.248	ME-Sad	0.746	ME-Fea	0.577	ME-Ang	0.458
	3	GT (DP)	0.139	GT (SC)	-0.008	GT (SC)	0.076	GT (SC)	0.288

Table IX lists comprehensive correlation analysis between some highly correlated social media signals and actual ground-truth cases, varying the horizons from -3 (three-month reflection) and 3 (three-month forecast).

References

[1] T. Tawichsri and B. Sangimnet, "Economic crisis and mental health," *PIER: aBRIDGEd*, vol. 8, 2021. [Online]. Available: https://www.pier.or.th/abridged/2021/08/

[2] M. Davlasheridze, S. J. Goetz, and Y. Han, "The effect of mental health on us county economic growth," *Review of Regional Studies*, vol. 48, no. 2, pp. 155–171, 6 2018.

[3] OECD, *A New Benchmark for Mental Health Systems*, 2021. [Online]. Available: https://www.oecd-ilibrary.org/content/publication/4ed890f6-en

[4] C. McClellan, M. M. Ali, R. Mutter, L. Kroutil, and J. Landwehr, "Using social media to monitor mental health discussions- evidence from twitter," *Journal of the American Medical Informatics Association*, vol. 24, no. 3, pp. 496–502, 2017.

[5] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *IEEE transactions on knowledge and data engineering*, vol. 26, no. 1, pp. 97–107, 2013.

[6] T. Bodnar, C. Tucker, K. Hopkinson, and S. G. Bilén, "Increasing the veracity of event detection on social media networks through user trust modeling," in *2014 IEEE International Conference on Big Data (Big Data)*. IEEE, 2014, pp. 636–643.

[7] D. Wardman, "Bringing big data to the enterprise–gaining new insight with big data capabilities," 2014.

[8] S. Tuarob, C. S. Tucker, M. Salathe, and N. Ram, "Discovering health-related knowledge in social media using ensembles of heterogeneous features," in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 2013, pp. 1685–1690.

[9] ——, "An ensemble heterogeneous classification methodology for discovering health-related knowledge in social media messages," *Journal of biomedical informatics*, vol. 49, pp. 255–268, 2014.

[10] ——, "Modeling individual-level infection dynamics using social network information," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 2015, pp. 1501–1510.

[11] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 851–860.

[12] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of computational science*, vol. 2, no. 1, pp. 1–8, 2011.

[13] S. Tuarob and C. S. Tucker, "Automated discovery of lead users and latent product features by mining large scale social media networks," *Journal of Mechanical Design*, vol. 137, no. 7, p. 071402, 2015.

[14] ——, "Quantifying product favorability and extracting notable product features using large scale social media data," *Journal of Computing and Information Science in Engineering*, vol. 15, no. 3, 2015.

[15] ——, "A product feature inference model for mining implicit customer preferences within large scale social media networks," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 57052. American Society of Mechanical Engineers, 2015, p. V01BT02A002.

[16] ——, "Discovering next generation product innovations by identifying lead user preferences expressed through large scale social media data," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 46292. American Society of Mechanical Engineers, 2014, p. V01BT02A008.

[17] ——, "Fad or here to stay: Predicting product market adoption and longevity using large scale, social media data," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 55867. American Society of Mechanical Engineers, 2013, p. V02BT02A012.

[18] ——, "Automated discovery of product preferences in ubiquitous social media data: A case study of automobile market," in *2016 International Computer Science and Engineering Conference (ICSEC)*. IEEE, 2016, pp. 1–6.

[19] P. Yin, Q. He, X. Liu, and W.-C. Lee, "It takes two to tango: Exploring social tie development with both online and offline interactions," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 9, no. 3, pp. 174–187, 2016.

[20] A. Sarker, K. O'connor, R. Ginn, M. Scotch, K. Smith, D. Malone, and G. Gonzalez, "Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from twitter," *Drug safety*, vol. 39, no. 3, pp. 231–240, 2016.

[21] C. Caragea, N. J. McNeese, A. R. Jaiswal, G. Traylor, H.-W. Kim, P. Mitra, D. Wu, A. H. Tapia, C. L. Giles, B. J. Jansen *et al.*, "Classifying text messages for the haiti earthquake." in *ISCRAM*. Citeseer, 2011.

[22] O. Gruebner, S. R. Lowe, M. Sykora, K. Shankardass, S. Subramanian, and S. Galea, "A novel surveillance approach for disaster mental health," *PLoS one*, vol. 12, no. 7, p. e0181233, 2017.

[23] H. Woo, Y. Cho, E. Shim, K. Lee, and G. Song, "Public trauma after the sewol ferry disaster: the role of social media in understanding the public mood," *International journal of environmental research and public health*, vol. 12, no. 9, pp. 10 974–10 983, 2015.

[24] A. H. Yazdavar, H. S. Al-Olimat, M. Ebrahimi, G. Bajaj, T. Banerjee, K. Thirunarayan, J. Pathak, and A. Sheth, "Semi-supervised approach to monitoring clinical depressive symptoms in social media," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 2017, pp. 1191–1198.

[25] D. L. Mowery, Y. A. Park, C. Bryan, and M. Conway, "Towards automatically classifying depressive symptoms from twitter data for population health," in *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES)*, 2016, pp. 182–191.

[26] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in twitter," in *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, 2014, pp. 51–60.

[27] R. D. Flores, "Do anti-immigrant laws shape public sentiment? a study of arizona's sb 1070 using twitter data," *American Journal of Sociology*, vol. 123, no. 2, pp. 333–384, 2017.

[28] Y. Gorodnichenko, T. Pham, and O. Talavera, "Social media, sentiment and public opinions: Evidence from# brexit and# uselection," *European Economic Review*, p. 103772, 2021.

[29] O. Metwally, S. Blumberg, U. Ladabaum, and S. R. Sinha, "Using social media to characterize public sentiment toward medical interventions commonly used for cancer screening: an observational study," *Journal of medical Internet research*, vol. 19, no. 6, p. e200, 2017.

[30] J. Prichard, P. Watters, T. Krone, C. Spiranovic, and H. Cockburn, "Social media sentiment analysis: A new empirical tool for assessing public opinion on crime?" *Current Issues in Criminal Justice*, vol. 27, no. 2, pp. 217–236, 2015.

[31] P. Grover, A. K. Kar, Y. K. Dwivedi, and M. Janssen, "The untold story of usa presidential elections in 2016-insights from twitter analytics," in *Conference on e-Business, e-Services and e-Society*. Springer, 2017, pp. 339–350.

[32] H. T. Le, G. Boynton, Y. Mejova, Z. Shafiq, and P. Srinivasan, "Revisiting the american voter on twitter," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 4507–4519.
[33] A. Darliansyah, H. M. Wandabwa, M. A. Naeem, F. Mirza, and R. Pears, "Long-term trends in public sentiment in indian demonetisation policy," in *International Conference on Intelligent Technologies and Applications*. Springer, 2018, pp. 65–75.
[34] H. Ishido, Y. Tashiro, and R. Liang, "Us president donald trump's twitter analysis and his trade policy agenda," *International Relations and Diplomacy*, vol. 6, no. 9, pp. 476–499, 2018.
[35] A. Mondschein, D. A. King, C. Hoehne, Z. Jiang, and M. Chester, "Using social media to evaluate associations between parking supply and parking sentiment," *Transportation Research Interdisciplinary Perspectives*, vol. 4, p. 100085, 2020.
[36] W. Chung and D. Zeng, "Social-media-based public policy informatics: Sentiment and network analyses of us immigration and border security," *Journal of the Association for Information Science and Technology*, vol. 67, no. 7, pp. 1588–1606, 2016.
[37] S. Chancellor, M. L. Birnbaum, E. D. Caine, V. M. Silenzio, and M. De Choudhury, "A taxonomy of ethical tensions in inferring mental health states from social media," in *Proceedings of the conference on fairness, accountability, and transparency*, 2019, pp. 79–88.
[38] A. Ceron and F. Negri, "The "social side" of public policy: Monitoring online public opinion and its mobilization during the policy cycle," *Policy & Internet*, vol. 8, no. 2, pp. 131–147, 2016.
[39] E. D'Avanzo, G. Pilato, and M. Lytras, "Using twitter sentiment and emotions analysis of google trends for decisions making," *Program*, 2017.
[40] D. Giannone, L. Reichlin, and D. Small, "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, vol. 55, no. 4, pp. 665–676, 2008.
[41] S. Tuarob, P. Wettayakorn, P. Phetchai, S. Traivijitkhun, S. Lim, T. Noraset, and T. Thaipisutikul, "Davis: a unified solution for data collection, analyzation, and visualization in real-time stock market prediction," *Financial Innovation*, vol. 7, no. 1, pp. 1–32, 2021.
[42] J. Ortega-Bastida, A. J. Gallego, J. R. Rico-Juan, and P. Albarrán, "A multimodal approach for regional gdp prediction using social media activity and historical information," *Applied Soft Computing*, vol. 111, p. 107693, 2021.
[43] P. Resnik, W. Armstrong, L. Claudino, T. Nguyen, V.-A. Nguyen, and J. Boyd-Graber, "Beyond lda: Exploring supervised topic modeling for depression-related language in twitter," *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2015.
[44] A. Benton, M. Mitchell, and D. Hovy, "Multitask learning for mental health conditions with limited social media data," *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 2017.
[45] P. Burnap, G. Colombo, R. Amery, A. Hodorog, and J. Scourfield, "Multi-class machine classification of suicide-related communication on twitter," *Online Social Networks and Media*, vol. 2, p. 32–44, 2017.
[46] D. Mowery, H. Smith, T. Cheney, G. Stoddard, G. Coppersmith, C. Bryan, and M. Conway, "Understanding depressive symptoms and psychosocial stressors on twitter: A corpus-based study," *Journal of Medical Internet Research*, vol. 19, no. 2, 2017.
[47] X. Chen, M. Sykora, T. Jackson, S. Elayan, and F. Munir, "Tweeting your mental health: An exploration of different classifiers and features with emotional signals in identifying mental health conditions," *The 51st Hawaii International Conference on System Sciences*, 2018.
[48] J. Weerasinghe, K. Morales, and R. Greenstadt, ""because... i was told... so much": Linguistic indicators of mental health status on twitter," *Proceedings on Privacy Enhancing Technologies*, vol. 2019, no. 4, p. 152–171, 2019.
[49] G. Coppersmith, R. Leary, P. Crutchley, and A. Fine, "Natural language processing of social media as screening for suicide risk," *Biomedical Informatics Insights*, vol. 10, p. 117822261879286, 2018.
[50] B. Verma, S. Gupta, and L. Goel, "A neural network based hybrid model for depression detection in twitter," *Communications in Computer and Information Science*, p. 164–175, 2020.
[51] Q. Cong, Z. Feng, F. Li, Y. Xiang, G. Rao, and C. Tao, "X-a-bilstm: A deep learning approach for depression detection in imbalanced data," *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018.
[52] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of suicide ideation in social media forums using deep learning," *Algorithms*, vol. 13, no. 1, p. 7, 2019.
[53] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, "Detecting depression and mental illness on social media: an integrative review," *Current Opinion in Behavioral Sciences*, vol. 18, pp. 43–49, 2017.
[54] A. Benton, M. Mitchell, and D. Hovy, "Multitask learning for mental health conditions with limited social media data," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Valencia, Spain: Association for Computational Linguistics, Apr. 2017, pp. 152–162.
[55] S. Tariq, N. Akhtar, H. Afzal, S. Khalid, M. R. Mufti, S. Hussain, A. Habib, and G. Ahmad, "A novel co-training-based approach for the classification of mental illnesses using social media posts," *IEEE Access*, vol. 7, pp. 166 165–166 172, 2019.
[56] S. Ghosh and T. Anwar, "Depression intensity estimation via social media: A deep learning approach," *IEEE Transactions on Computational Social Systems*, pp. 1–10, 2021.
[57] J. Zhou, H. Zogan, S. Yang, S. Jameel, G. Xu, and F. Chen, "Detecting community depression dynamics due to covid-19 pandemic in australia," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, pp. 982–991, 2021.
[58] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, no. 7232, pp. 1012–1014, Feb. 2009, bandiera_abtest: a Cg_type: Nature Research Journals Number: 7232 Primary_atype: Research Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/nature07634
[59] N. Tefft, "Insights on unemployment, unemployment insurance, and mental health," *Journal of Health Economics*, vol. 30, no. 2, pp. 258–264, Mar. 2011.
[60] Q. Lhoest, A. V. del Moral, Y. Jernite, A. Thakur, P. von Platen, S. Patil, J. Chaumond, M. Drame, J. Plu, L. Tunstall, J. Davison, M. vSavsko, G. Chhablani, B. Malik, S. Brandeis, T. L. Scao, V. Sanh, C. Xu, N. Patry, A. McMillan-Major, P. Schmid, S. Gugger, C. Delangue, T. Matussiere, L. Debut, S. Bekman, P. Cistac, T. Goehringer, V. Mustar, F. Lagunas, A. M. Rush, and T. Wolf, "Datasets: A community library for natural language processing," *ArXiv*, vol. abs/2109.02846, 2021.
[61] G. de Melo and S. Siersdorfer, "Multilingual text classification using ontologies," in *Advances in Information Retrieval*, G. Amati, C. Carpineto, and G. Romano, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 541–548.

[62] R. Mihalcea, C. Banea, and J. Wiebe, "Learning multilingual subjective language via cross-lingual projections," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Prague, Czech Republic: Association for Computational Linguistics, Jun. 2007, pp. 976–983.

[63] C. Banea, R. Mihalcea, J. Wiebe, and S. Hassan, "Multilingual subjectivity analysis using machine translation," in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. Honolulu, Hawaii: Association for Computational Linguistics, Oct. 2008, pp. 127–135.

[64] X. Wan, "Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis," in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. Honolulu, Hawaii: Association for Computational Linguistics, Oct. 2008, pp. 553–561.

[65] M. Salameh, S. Mohammad, and S. Kiritchenko, "Sentiment after translation: A case-study on Arabic social media posts," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Denver, Colorado: Association for Computational Linguistics, May–Jun. 2015, pp. 767–777. [Online]. Available: https://aclanthology.org/N15-1078

[66] S. M. Mohammad, M. Salameh, and S. Kiritchenko, "How translation alters sentiment," *J. Artif. Int. Res.*, vol. 55, no. 1, p. 95–130, Jan. 2016.

[67] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[68] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov, "Fasttext. zip: Compressing text classification models," *arXiv preprint arXiv:1612.03651*, 2016.

[69] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 2227–2237. [Online]. Available: https://aclanthology.org/N18-1202

[70] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 328–339. [Online]. Available: https://aclanthology.org/P18-1031

[71] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423

[72] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[73] Y. Yang, G. Hernandez Abrego, S. Yuan, M. Guo, Q. Shen, D. Cer, Y.-h. Sung, B. Strope, and R. Kurzweil, "Improving multilingual sentence embedding using bi-directional dual encoder with additive margin softmax," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 7 2019, pp. 5370–5378. [Online]. Available: https://doi.org/10.24963/ijcai.2019/746

[74] T. Pires, E. Schlinger, and D. Garrette, "How multilingual is multilingual BERT?" in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 4996–5001. [Online]. Available: https://aclanthology.org/P19-1493

[75] A. Conneau, R. Rinott, G. Lample, A. Williams, S. Bowman, H. Schwenk, and V. Stoyanov, "XNLI: Evaluating cross-lingual sentence representations," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 2475–2485.

[76] A. Conneau and G. Lample, "Cross-lingual language model pretraining," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019.

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

[78] S. Renjit and S. M. Idicula, "CUSATNLP@DravidianLangTech-EACL2021:language agnostic classification of offensive content in tweets," in *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Kyiv: Association for Computational Linguistics, Apr. 2021, pp. 236–242. [Online]. Available: https://aclanthology.org/2021.dravidianlangtech-1.32

[79] O. Gencoglu, "Large-scale, language-agnostic discourse classification of tweets during covid-19," *Machine Learning and Knowledge Extraction*, vol. 2, no. 4, pp. 603–616, 2020. [Online]. Available: https://www.mdpi.com/2504-4990/2/4/32

[80] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A dataset of fine-grained emotions," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, Jul. 2020, pp. 4040–4054. [Online]. Available: https://aclanthology.org/2020.acl-main.372

[81] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[82] H.-C. Shing, S. Nair, A. Zirikly, M. Friedenberg, H. Daumé III, and P. Resnik, "Expert, crowdsourced, and machine assessment of suicide risk via online postings," in *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*. New Orleans, LA: Association for Computational Linguistics, Jun. 2018, pp. 25–36. [Online]. Available: https://aclanthology.org/W18-0603

[83] A. Zirikly, P. Resnik, Ö. Uzuner, and K. Hollingshead, "CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts," in *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 24–33. [Online]. Available: https://aclanthology.org/W19-3003

[84] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to information retrieval*. Cambridge University Press Cambridge, 2008, vol. 39.

[85] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *ADVANCES IN LARGE MARGIN CLASSIFIERS*. MIT Press, 1999, pp. 61–74.

[86] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
[87] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional lstm and other neural network architectures,” *Neural networks*, vol. 18, no. 5-6, pp. 602–610, 2005.
[88] I. Loshchilov and F. Hutter, “Fixing weight decay regularization in adam,” *CoRR*, vol. abs/1711.05101, 2017. [Online]. Available: http://arxiv.org/abs/1711.05101
[89] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: Synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, p. 321–357, Jun 2002. [Online]. Available: http://dx.doi.org/10.1613/jair.953
[90] VISTEC-depa, “English-Thai Machine Translation Models,” https://airesearch.in.th/releases/machine-translation-models/, 2020, [Online; accessed 25-December-2020].
[91] E. Ansari, A. Axelrod, N. Bach, O. Bojar, R. Cattoni, F. Dalvi, N. Durrani, M. Federico, C. Federmann, J. Gu, F. Huang, K. Knight, X. Ma, A. Nagesh, M. Negri, J. Niehues, J. Pino, E. Salesky, X. Shi, S. Stüker, M. Turchi, A. Waibel, and C. Wang, “FINDINGS OF THE IWSLT 2020 EVALUATION CAMPAIGN,” in *Proceedings of the 17th International Conference on Spoken Language Translation*. Online: Association for Computational Linguistics, Jul. 2020, pp. 1–34. [Online]. Available: https://aclanthology.org/2020.iwslt-1.1
[92] N. Avery, J. Ghandi, and J. Keating, “The ‘dr google’phenomenon–missed appendicitis,” *NZ Med J*, vol. 125, no. 1367, pp. 135–137, 2012.