

THE NEW PERCEPTION OF HEALTH INSURANCE PACKAGE HARNESSING SOCIAL MEDIA FOOTPRINT

Nattarika Poolperm, *Theeraya Mayakul, Supaporn Kiattisin and
Somkiat Wattanasirichaigoon

Information Technology Management Division, Faculty of Engineering Mahidol University of Thailand, Thailand

ABSTRACT

Social media is an incredible tool for individuals to communicate people's interests and opinions that represent their emotions, thoughts, and sentiments. Social media data has been used to detect mental health in several studies. Health insurance is always evolving and looking for improvement. This study proposes a new concept in health insurance to adopt social media data as one of the critical factors for designing and pricing products. The data given as comments, hashtag, keywords, and time spent are used to classify the risk of psychological disorder, the mental health problems are covered. The study suggests the machine learning technique, especially Natural Language Processing, as an efficient method. In addition, social media data and health information are considered sensitive personal data; the framework must comply with GDPR, HIPAA, and Thai PDPA regulations. Ten experts in the health insurance industry in Thailand are invited to participate in the survey evaluation. The item-content validity index (I-CVI) and Kappa statistic were assessed by six domain experts and four general experts, each with ten years of experience. The questionnaires were identified for experts' agreement, satisfaction, and acceptance. The I-CVI range result was very high (0.70 to 1), and the modified Kappa agreement among the experts validated as an excellence ($k > 0.70$). Overall, the feedback was positive. Therefore, the framework was satisfied and accepted to implement.

Keywords: health insurance, social media, mental illness, natural language processing

1. INTRODUCTION

The World Health Organization (WHO) has analyzed that in today's world, traumatic mental disorders have escalated in all countries. The examples of the mental illnesses are depression, bipolar disorder, schizophrenia and other psychoses, dementia, anxiety, and

developmental disorders such as autism[1]. In the digital era, numerous studies illustrate the connection between social media and mental and physical health problems. It is critical to discover mental health problems at the early stage as it is beneficial in preventing additional harm. Moreover, health insurance is able to precisely design a health insurance plan because the treatment of mental health problems is pharmacotherapy, which is expensive [2].

The essential act of health insurance is to delegate part of the financial burden of medical bills and provide a risk transfer technique in exchange for an insurance premium. Thus, determining the appropriate insurance product is one of the most significant responsibilities in the insurance industry [3]. The various factors that are normally utilized for pricing and designing health insurance packages are demographics, smoking, children, BMI, health condition, geographic, occupation, and lifestyle. These might not cover all present-day. As the goal of WHO, patients with mental disorders must have access to the precise treatment and must be covered by health insurance [1].

Therefore, the purpose of this study is to utilize social media data as one of the key factors in delivering an accurate insurance package. To protect the health related data, we adopt the privacy by design principle.

The remaining paper is organized as follows: Section II provides a forceful overview of the health insurance function, the connection between health conditions with social media, and the privacy regulations. Further, the machine learning technique; Natural Languages Processing (NLP) is presented in Section II. Section III presents the methodology of the proposed conceptual framework. Section IV illustrates the survey evaluation. The results and discussion are in section V and VI, followed by the conclusion and prospects of harnessing social media data for pricing and designing health insurance package.

2. LITERATURE AND RELATED WORK

The literature review aims to develop the concept of social media data on health status. The opportunity, strategy, and process in the insurance business are identified to develop these innovation initiatives. In addition, the emerging technology and data management framework may incorporate to ensure privacy and efficiency. We found several research studies concerning the use of social media with technology that can detect mental health problems. These are as follows;

Manuscript received on September 17, 2022, revised on November 03, 2022, accepted on November 14, 2022.

*Corresponding author Email: Theeraya.may@mahidol.ac.th
Division of Information Technology Management, Faculty of Engineering, Mahidol University, Thailand.

A. Health Insurance

Insurance plays an important role as a tool for minimizing financial risk [4]. Medical insurance is a means of transferring the risk of illness or injury. Purchasing medical insurance helps the insured to delegate part of the financial burden of medical bills to the insurer, allowing the insured to make better informed decisions about medicines and treatments they receive [5].

People always count on health coverage to minimize the expenses of medical care in the event of an accident or illness. Healthcare services are often out of reach for those who cannot afford to pay for them out-of-pocket or whose financial situation makes it impossible. However, a few essential medical treatments are covered by public and commercial health insurance in low- and middle-income nations, exposing these citizens to significant financial risk. The health insurance system does not cover many conditions and some novel treatments [6].

The process of health insurance initiation composes of the pocket design, finance management, contract agreement, and reimbursement. The descriptions are as follows;

Insurance Policy

Policy administration is one of the most critical responsibilities in insurance industry. Insurers and insureds sign a contract known as an insurance policy, laying out the terms of the agreement. Additionally, plans place duties on the insured in addition to providing protection. Declarations, insuring agreements, definitions, exclusions and conditions are the five main components of every insurance policy [7].

Product Development

Product development is the core of all insurance activity. This process is the involvement of underwriters and the actuary. The actuary and underwriter work together to assess the risk profile of each risk type that the insurance company is willing to accept. Underwriting criteria, a unique application form, and relevant exclusions will be used to evaluate the risk profile. The actuaries may then price the items after they have the risk profile [8].

The provider has delivered pockets of innovation to achieve the new customer expectation with value interest rates. The high competition would be added by an insurtech.

Role of Actuary

The actuary is one of the professions featured in insurance businesses, where they are in charge of pricing products, calculating obligations, managing risk, making investments, and determining if a company's capital is adequate. On the other hand, professional actuaries are mathematicians who use their expertise to assess the likelihood and danger of future events. These abilities are

used to predict how certain occurrences will affect a company's bottom line and its clients' bottom lines [9]. Sophia Dyson et al [10] describe that an actuary in health insurance performs a wide range of duties, including:

- Estimating outstanding claims
- Pricing and designing products
- Determining the fund's solvency and capital adequacy position
- Projecting business indicators for a variety of purposes including budgeting, strategic planning and corporate mergers
- Analyzing utilization and claims cost experience
- Negotiating with hospital negotiations and calculating second tier benefit schedules.

According to relevant literature research, the cost of health insurance premiums is based on seven factors including age, gender, BMI, number of children, smoking status, region, and charge level of individuals [10]. The use of prediction analysis has improved the accuracy of health insurers' premium pricing and established an individual health insurance plan type [11]. In non-life insurance, there are often two distinct roles. Two types of actuaries are 1) the pricing actuary who creates and rates new insurance products and 2) the reserving actuary who forecasts insurance claim cash flows. These forecasts are utilized in insurance accounting, risk management, and product creation. Predicting the best outcomes is what actuaries are primarily concerned. Yet, they prefer a degree of model complexity that can be readily articulated to both management and customers [12].

Underwriting

Underwriting serves the function of identifying and classifying potential risks to a business. However, risks are not evaluated on an individual basis. As a result, individuals are classified into risk categories (class rating) and pay premiums based on the average risk profile of their class, rather than relying solely on demographic, personal history, and genetic data. An insurance underwriter examines, accepts, or rejects insurance risks and then categorizes them in order to charge the appropriate premium for each. It is the job of underwriters to distribute risk among a group of insureds in a way that is fair to the insureds and profitable for the insurer [13]. Avoiding the unfairness of a premium, the actuary and underwriter have to work together to calculate the premium and design a product based on the risk of the insured.

B. Social Media

Presently, it is undeniable that social media websites and interacting with others through online platform have become the inseparable tool for social gathering [14]. Social media is a computer-based technology[15], which allows users and communities to gather and communicate by establishing virtual networks and communities; to share information, ideas, personal messages, images, and other content; and, in some cases, to collaborate with

others in real time. Social media are also referred to as “Web 2.0” or “social networking.” [16]

Social media contain a lot of data including the users’ attitude, interest, content and news. Some people say that data is the new gold especially in the digital era, because it represents insights into owner’s behavior and life style.

Social Media and Health Problem

Although social media is convenient, it can be a double-edge sword. The research of Chloe Berryman et al.[14] stated that the impact of social media on mental health are currently presented as a significant problem.

Prior research studies have found how to detect mental illness on social media posts. In some cases, researchers used keywords on Twitter associated with mental illness, such as “depressive” and “successful suicide,” to find potential candidates for the study. Other terms that could indicate mental illness symptoms include “distress,” “gloom,” “blue,” “empty,” “sad,” “hate,” “kill,” “do not want to live anymore,” “ashamed of myself,” and so on. Moreover, their systematic review illustrates that some studies have created an experimental dataset. The “Reddit Self-reported Depression Diagnosis (RSDD) dataset” is an experimental dataset that comprises roughly 9K diagnosed users and over 100K control users. a total of 49,580 depressed and 481,873 non-depressed postings were gathered. Each of the following clinical subreddits had 56,009 posts in total. Regression models were trained by using unigram word vectors to categorize the post to one of the four listed mental illnesses: ADHD, Anxiety, Bipolar, and Depression [17].

Gautami Tripathi and Mohd Abdul Ahad [18] presented that most people spend significant time on the internet and social networking due to technical advancements. Their study showed that prolonged hours spent on social media sites and staying online have not only caused severe damage to the mental health of individuals but also affected on other conditions such as social relationships and behavior.

In addition, Kenneth C. Laudon and Jane P. Laudon [19] explained that burning eyes or Computer Vision syndrome induces almost 90 percent of people who spend more than three hours daily on a computer or other electronic devices. Its symptoms are normally impermanent, including headaches, blurred vision, dry and irritated eyes.

In addition, this behavior is making users incapable of thinking beyond their virtual life, causing depression, hyperactivity, mood swings, lack of reasoning and analysis, insomnia [20] ADHD (attention deficit hyperactivity disorder), anxiety, bipolar disorder [21], as well as Post Traumatic Stress disorder (PTSD) [22].

In sum, social media activities and behavior can assist to predict health problems, including physical and mental disorders.

C. Personal Health Information

Personal health information defines as medical history, laboratory results, demographic information, mental and physical health condition, insurance information, and any form of information that a healthcare professional collects to identify an individual and determine appropriate care [23].

GDPR in Healthcare

According to Recital 35 in the GDPR, defined personal health information is defined as “*Personal data concerning health should include all data pertaining to the health status of a data subject which reveal information relating to the past, current or future physical or mental health status of the data subject.*” [24] Previous studies have discussed that “Sensitive” data, such as medical and genetic information, are subject to strict regulations. Before processing any data, several rules and regulations must be applied. The condition stipulates that the data subject's “explicit” permission must have been provided if secure data subjects to patients who are unable to offer their permission, such as an unconscious patient's medical emergency. A physician enables to requires data from other providers to deliver quality treatment. However, public safety or health security is a must in exchanging data among providers to prevent cross-border health hazards [25].

HIPAA in Healthcare

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) prohibits unauthorized access to medical records and can constitutes a HIPAA breach. Healthcare providers are required to maintain the confidentiality of patients' medical records under HIPAA.

HIPAA governs the use and disclosure of protected health information (PHI) in the United States. PHI, according to HIPAA, is information concerning mental or physical health of a person. HIPAA only applies to a certain group of organizations: health care plans and health care payment systems. HIPAA protects organizations and groups in order to offer PHI security and privacy. Unless an exemption applies, organizations and business associates cannot disclose or use PHI without prior patient consent[25]. In accordance with HIPAA laws, PHI is defined as “*individually identifiable health information*”[26]. “Individually identifiable health information is described as data, including demographic data, about an individual's past, present, or future physical or mental health or condition, the provision of health care to the individual, or the past, present, or future payment for the provision of health care to the individual that identifies the individual or has a reasonable basis to believe can be used to identify the individual. Individually identifiable health information has a variety of standard identifiers such as name, address, birth date, and the social security number [27].”

Thailand PDPA in Healthcare

Thailand's Personal Data Protection Act (PDPA) is very similar to the European GDPR. It clearly defines that data containing personal identifiers must be restricted in the use of the data because of privacy concerns [28].

In section 23 of PDPA regulation, collecting personal data, the Data Controller shall inform data subject regarding; the purpose of the collection, notification of the case, the retention period, the categories of persons or entities to whom the collected data maybe disclosed, information contact of the data controller, and the rights of the data subject [29]. According to section no. 24 in PDPA, it has described that the data controller shall not collect personal data without consent of the data subject, unless it is for preventing or suppressing a danger to a person's life, body or health [29].

Social Media Data and Privacy

Social media data can monitor the behavior of user, especially when surfing the internet, such as visited websites or the location of the checked-in feature. In addition, social media gathers information that users provide, for example, marital status, age, employment, places, etc. For population-level social media research, privacy has been recognized as an essential moral dilemma [30]. Social media websites not only connect people but also are excellent data collectors, gathering a broad array of information ranging from non-sensitive to highly sensitive information. Although many types of data may be non-sensitive on their own, combining different types of data may lead to insights into sensitive health issues.

It is necessary for a data subject to be aware of the data being collected or used to determine whether the information is sensitive. The relevant information should respond to needs, as well as the data quality must be accurate and complete[31].

There are two dimensions of informed consent in social media: 1) the informed consent that users give to the service providers, and 2) the other concerns the specific use of personal data for research purposes [32].

Hence, the purpose of gathering data on social media must be permitted by the informed consent of users.

D. Machine Learning: Natural Languages Processing (NLP)

Natural Language Processing (NLP) is examined with the computational processing of human natural language in the form of texts, such as post and tweet, whereas Machine Learning (ML) is dealt with the development of data-driven computational algorithms that can "learn" from data in order to recognize patterns and make predictions. According to the study by Mike Conway and Daniel O'Connor [30], psychological health has been a research topic for NLP researchers since the discipline's inception. NLP can determine sentiment, affect-related states, basic emotions such as anger, fear,

or happiness, as well as mental health illnesses like post-traumatic stress disorder (PTSD) and depression [22].

Furthermore, it can provide insights into mental health and emotional state of a person in addition to their ability to communicate in a variety of ways, including the use of narrative, subjective or structured forms of speech as well as their socio-economic status and living situations [33].

There are tremendous prospects to improve evidence-based decision making in public health through NLP. The use of NLP in a broader range of applications will lead to more efficient surveillance systems that can identify diseases and at-risk conditions in real time [34]. In 2016, Mike Conway and Daniel O'Conno's [30] discovered that recent works have been using NLP and ML methods to assess suicide risk in pediatric populations based on writing samples and to predict depression severity and optimal treatment. In accordance with the advancements in NLP, there is a long tradition in psychology of using carefully developed and validated lexicons organized into various categories (e.g., anxiety, insight, achievement) to score texts based on the existence or absence of mental terms [30]. Another example is the work done by Benjamin L. Cook et al [35] that compared NLP-based models with logistic regression prediction models to predict suicidal ideation and heightened psychiatric symptoms using unstructured questions like "how do you feel today?" and using structured data example sleep pattern and well-being. The findings indicated that models built on NLP could generate pretty high prediction values based merely on replies to a basic general mood inquiry [35].

3. METHODOLOGY

We have designed a conceptual framework which is based on a systematic review. Besides, the contribution of this paper is to develop a business framework in health insurance to provide an accurate estimate of the future premium and the health insurance package. Therefore, to achieve that we have summarized the analyzing mental health problem that is detected from social media. In Table 1 we have generated the data for the purpose to collect in the framework; comments including hashtag or keywords and behavior such as time spending.

According to the literature review, mental and physical health problems are one of the disadvantages caused by prolonged hours of staying online on social media. On the other hand, mental disorders can also be detected from post and comments. Common mental illnesses usually include depression, bipolar, insomnia, and anxiety, shown in Figure 1.

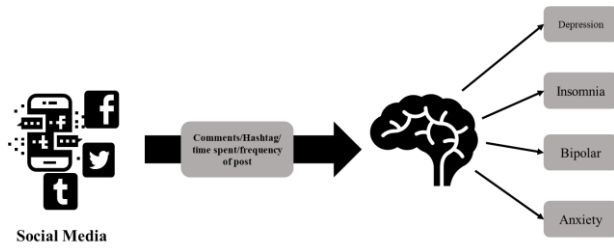


Figure 1 Common mental illnesses

Table 1 Type of data

Data	Type of data	Examples
Comments	Unstructured	Hashtag or Keywords
Behavior	Unstructured	Time spending

Data privacy is a critical issue that needs to be addressed. We adopted the privacy regulations, including GDPR, HIPAA, and Thailand PDPA to develop the framework for all platforms. Thus, the term of condition and informed consent must be provided for an insured agreement to protect against unawareness. An insured must understand and should accept. The detail must cover in the following 7 keys:

- The insured acknowledges that the insurer has the right to gather data and information from his/her social media websites to design the health insurance package and for marketing research.
- The insurer has the right to deny if the insured provides false statement.
- The insured must notify the insurer of any changes in his/her status or personal information within 30 days.
- The insurer is not allowed to share any social media data of the insured to any third-party without authorization.
- Medical data including diagnostic of mental health, which has been detected from social

media, will be disclosed or anonymized to the partner hospital.

- The insured can withdraw consent at any time.
- All sensitive data that the insurer collects will be concealed

Conceptual Framework

Based on the insurance business, social media, technology and governance, we have summarized the process in Figure 2. The process starts with collecting social media data and medical records. Then, all unstructured data are analyzed and at the same time, the insurer verifies the medical and claim history. Finally, a health insurance package is generated.

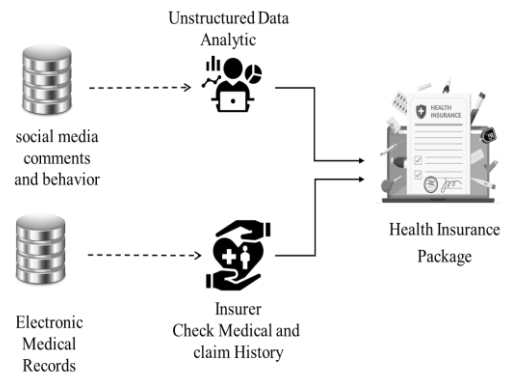


Figure 2 Health insurance package process

The personal data must be used under limitations. Hence, data will not be disclosed, made available, or used for other purposes besides designing health insurance packages and predicting premiums. To follow up on the treatment on a mobile application, the insured must first enter his/her username and password. Moreover, after signing in, the insured must first identify by receiving the OTP via SMS and then using the fingerprint biometric security system. Only the electronic fingerprint biometric is accepted to update data and to edit any information on the application. The procedure of data gathering is separated into five steps:

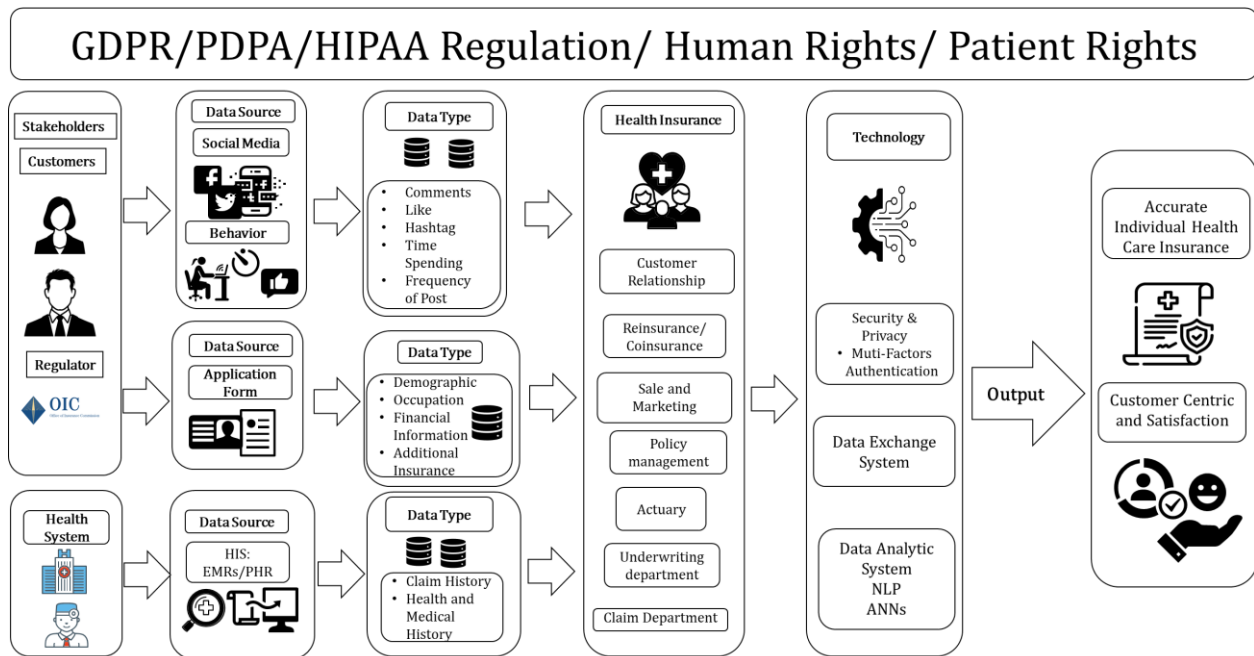


Figure 3 Conceptual Framework

1. The collected social media data is processed into the health insurance system. The system is encrypted and limited to the pocket and product management, the claim department, the marketing team for marketing research, and the partner hospital. The data storage period depends on the insurance plan and the customer’s decision. After the insurance policy is terminated, all the social media data will remain in the system for five years, and the insured acknowledges this by informed consent. The informed consent is accepted. Social media comments and usage behavior are collected as key data.
2. The technical team (Underwriter and Actuary) analyzes data using machine learning to predict the risk of mental and physical disorders.
3. The information regarding the medical and claim records are verified.
4. The actuarial department and underwriting team design the premium and personalize the insurance package from the social behavior and underlying medical condition.
5. The accurate health insurance package, which covers the treatment cost, is provided as shown in figure 3.

Figure 3 shows the conceptual framework for developing the health insurance business. There are three external key stakeholders, including customer (insured), regulator (Office of Insurance Commission, OIC), and health provider (hospital). The data is categorized into three types. The framework identifies the data like comments, hashtags, keywords, and time spent to classify and evaluate the risk of mental disorder. Other data is collected from application forms and medical health records from hospitals. Then the health insurance

company utilizes the data from the mentioned resources to provide a health insurance package. The product development team is the most important step, especially actuary and underwriting. The data stakeholders are involved in customer relationship, claim, package and premium analysis. In the framework, the researcher recommends the three strategy technologies ;1) multi-factor authentication for security and privacy; 2) data exchange system ; 3) data analytic system to enable the Intelligence process. The NLP and Artificial Neural Networks (ANNs) are considered algorithm for health risk and disease prediction. Therefore, the individual health insurance is generated. The feedback data given as the package satisfaction and engagement are collected to strengthen the capability .

4. EVALUATION

The framework has been designed. We have conducted an expert panel to evaluate the framework. The survey questionnaires were initiated to measure the satisfaction and acceptance of this novel concept. The questionnaire focused on satisfaction, acceptability, efficacy, and applicability [36]. The responses were given on a four-point scale—completely dissatisfied, dissatisfied, satisfied, and strongly satisfied—to avoid the neutral point.

We have asked the group of experts to give their viewpoints on the questionnaires for the satisfaction and acceptance of using social media data in designing insurance plans. According to the suggestion of Shrotryia Vijay Kumar et al., the Content Validity Index (CVI) is a suitable statistic in social, administrative, and health science research to evaluate the viewpoint of experts [37]. We used CVI to calculate for all individual items. CVI is

a technique for determining context validity. The following is a five-step procedure:

- 1) Define idea, domain, and item.
- 2) Create a CVI form (Evaluation dimension)
- 3) Specify a qualification of experts.
- 4) Gather information from each expert.
- 5) Evaluate and discuss.

Enas Almanasreh et al. [38] suggested that ten experts can reduce the probability of chance agreements and may better inform instrument development.

Therefore, a sample of 10 experts have been selected. In details, six experts were set as the domain experts: three actuaries and three underwriters with ten years experiences in package designing and premium prediction as shows in Figure 4.

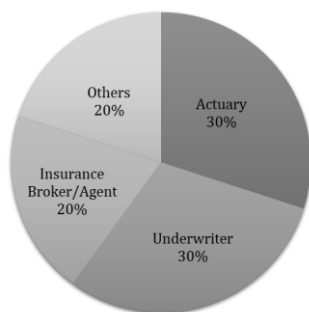


Figure 4 Evaluation Expert Roles

There are two types of CVI which are I-CVI and S-CVI. In the initial phase, I-CVI was calculated for each item, followed by S-CVI/UA and S-CVI/Ave, respectively. S-CVI is scale validity for overall agreement. Each item is separated into four scales according to the Likert scale: 4 = highly relevant, 3 = rather relevant, 2 = somewhat relevant, and 1 = not relevant in the I-CVI calculation, which is to identify the association per item. Experts will examine each material based on opinion and experience in a validity test. The I-CVI computation divides the total number of experts by the number of experts assigned to levels 3-4. The I-CVI calculation results have the following criteria: a range of 0.00-1.00. If there are between five and ten experts, the acceptable I-CVI value should not be less than 0.78; if there are fewer than five experts, the accepted I-CVI value should be 1.00. It must be updated again if the resulting condition is less than the set value [38].

A modified kappa k^* method for content validity adjusts each I-CVI for chance agreement. The recommended kappa value is above 0.60 and is substantial, while some research suggests values of 0.75 or more to be excellent. The value of each k^* is at an excellent level when $k > 0.74$, a good value (0.60-0.74) and a fair value (0.40-0.59), respectively [38].

The probability of chance agreement (P_c) is estimated by k^* equation as follows:

$$P_c = (N! \div A! (N-A)!)0.5N$$

Where, N= number of experts; and A= number of experts giving the agreement.

The Kappa modified (k^*) may be determined using the value of P_c and I-CVI after determining the I-CVI for all items using the following equation:

$$k^* = I-CVI - P_c / 1 - P_c$$

To promote the framework as a new digital policy, it needs cooperation among the business, technology, and governance. Hence, we adopted The Organization for Economic Co-operation and Development (OECD-DAC) evaluation framework for qualitative measurement. It comprises six criteria: relevance, coherence, effectiveness, efficiency, impact, and sustainability. Therefore, the evaluation questions are based on the OECD-DAC Criteria. However, not all six are suitable for our study because it is a pilot survey. We selected three criteria, namely efficiency, impact, and sustainability to test the concept before implementation [39].

5. RESULTS

The I-CVI for all items results from 0.70 to 1. Kappa statistic has been validated as an excellent agreement ($k > 0.70$). The Content validity index (CVI) of all experts' agreement from the above result are illustrated in table 3.

The answer to all eleven questions can be discussed as follows;

Question 1: Overall, from your experience, how would you rate this conceptual framework?

All ten experts replied that they thought this framework is useful and they gave a high score with this new concept. However, one expert suggested that the framework principles must be clear before execution.

Question 2: In your opinion, how are you satisfied with collecting social media comments to predict premium amount?

This question was provided to ask whether the social media data can be helpful in predicting premium rate and all of them were satisfied to collect the data from social media. They agreed that social media data could be a valuable resource.

Question 3: Do you think this conceptual framework can benefit the insurance business?

All experts gave a positive answer and thought this framework could help the mental health patients.

Question 4: Do you think this conceptual framework is useful to design a health insurance package?

The experts thought this framework could be applied to design a health insurance package.

Question 5: If this conceptual framework has been implemented. Do you agree or disagree that this framework can deliver the accurate premium amount?

The experts had varied responses. There were eight experts agreed and two experts disagreed. It could be argued that there should be more than one variable to calculate the accurate future premium. Moreover, the premium could be reduced if the insured has recovered from their mental problem.

Question 6: Do you agree or disagree that mental health can be detected from social media comments?

Table 3 CVI statistic result

Item/Expert	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Expert 8	Expert 9	Expert 10	Number of Agreement	I-CVI	Pc	Kappa Statistic
Q1	4	4	3	4	3	4	3	4	4	3	10	1.00	0.000537109	1.00
Q2	4	3	4	3	3	3	3	3	4	3	10	1.00	0.000537109	1.00
Q3	4	3	3	3	3	3	3	4	4	3	10	1.00	0.000537109	1.00
Q4	4	3	4	4	3	3	3	4	4	2	9	0.90	0.001193576	0.90
Q5	4	3	3	1	3	4	3	3	3	1	8	0.80	0.00201416	0.80
Q6	4	3	4	3	3	4	3	3	3	1	9	0.90	0.001193576	0.90
Q7	4	3	3	3	3	4	3	4	4	3	10	1.00	0.000537109	1.00
Q8	4	3	3	3	3	4	3	4	4	1	9	0.90	0.001193576	0.90
Q9	4	3	3	3	2	4	3	4	2	2	7	0.70	0.003069196	0.70
Q10	4	3	3	4	3	4	3	3	4	1	9	0.90	0.001193576	0.90
Q11	4	3	3	4	3	4	3	3	4	1	9	0.90	0.001193576	0.90
											Mean I-CVI	0.90		
											Sum of I-CVI	10.9		
Number of agreement	11	11	11	10	11	11	11	11	10	4	Ave-CVI ²	0.99		
Proportion of agreement	1	1	1	0.9	1	1	1	1	0.9	0.3	UA-CVI ³	0.36		

Ave- proportion of agreement across experts (Mean expert proportion) 0.91

Actuaries had an opinion regarding the social media comment that only comments could not be able to tell. It should have a second opinion or a psychologist's reference to confirm the diagnosis. Another expert explained that it would depend on the collected data's timeline.

Question 7: If this conceptual framework has been implemented. Do you agree with sharing the social media data with partner hospitals to provide the right treatment?

All experts agreed that sharing this data with partner hospitals would be very useful.

Question 8: Do you agree or disagree that this framework has important benefits in society?

90% of the experts agreed that this framework is beneficial to our digital society.

Question 9: Do you agree or disagree that the comment of "I feel lonely" could be a sign of mental illness?

Three experts disagreed concerning this phrase. It is insufficient to be a sign of mental sickness. They gave practical advice that we have to look for the frequency of how often the person posted it online. In addition, it must also be evaluated together with other variables. *Question 10: How are you satisfied to detect mental health from social media by using Natural Language Processing (NLP) Model?*

Most of the experts were satisfied and thought that Natural Language Processing (NLP) Model could assist them to identify mental health problems from messages on social media.

Question 11: How are you satisfied to generate the accurate premium amount by using Artificial Neural Networks Model (ANNs)?

Nine experts replied that they felt this model could be the tool to support the NLP model in forecasting future premiums.

6. DISCUSSION

Overall, the responses of the experts were positive. The ten experts appeared more concerned with detecting

mental health from social media data and gave suggestions about the implementation.

Although the framework received constructive feedback, the verification stage should be added in advance to detect fake social media data and prevent bias. It can be conducted manually at the beginning. Once there is sufficient data, AI can be replaced in accordance with ethics and responsibility. However, all the information on social media cannot explain everything, especially the financial statements of the insured. It needs multiple sources of data to ensure and optimize the data quality in terms of trust and integrity. Thus, information concerning the source of income should be provided in the application form.

According to comments on social media, most experts have agreed that social media data can be a useful resource for detecting mental health problems at the early stage. Even though some of them have thought that other variables are still needed to support before a diagnosis because mental illnesses have complex symptoms.

The participants, who are both actuaries and underwriters, have provided opinions: they have thought that this framework can support other variables to provide an accurate health insurance package and premium, which can be helpful for people with a mental health condition.

7. CONCLUSION

This research has successfully developed a framework for designing a health insurance package for mental health patients. This framework was designed to provide accurate health insurance premiums and packages by incorporating social media behavior. The data given as comments, hashtag, keywords, and time spent are factors in classifying the risk of mental disorder. In addition, these innovations could be used to prevent the harmful and recurrence of mental illnesses from social media.

The CVI experts' agreement result has 0.90. This result is acceptable. Therefore, harnessing social media

data can be a key factor for pricing and designing health insurance packages in the future. The predictive data model's technology is accepted and is a potential advantage. However, it is a challenge to implement the use of social media data as a diagnosis tool for mental health problems. For social media research, the issue of privacy has been a key ethical consideration. Hence, our framework complies with GDPR, HIPAA, and Thailand PDPA regulations. With the survey results, the experts are satisfied and accept this framework. They have been strongly agreed to implement in the future.

REFERENCES

- [1] "Mental disorders," *World Health Organization*, 2019. <https://www.who.int/news-room/factsheets/detail/mental-disorders> (accessed Nov. 28, 2019).
- [2] M. Sado *et al.*, "Cost-effectiveness analyses of augmented cognitive behavioral therapy for pharmacotherapy-resistant depression at secondary mental health care settings," *Psychiatry Clin. Neurosci.*, 2021, doi: 10.1111/pcn.13298.
- [3] K. M. Sakthivel and C. S. Rajitha, "Artificial Intelligence for Estimation of Future Claim Frequency in Non-Life Insurance," *Glob. J. Pure Appl. Math.*, vol. 13, no. 6, pp. 1701–1710, 2017, [Online]. Available: <http://www.ripublication.com>.
- [4] B. D. Sommers, A. A. Gawande, and K. Baicker, "Health Insurance Coverage and Health — What the Recent Evidence Tells Us," *N. Engl. J. Med.*, vol. 377, no. 6, pp. 586–593, 2017, doi: 10.1056/nejmsb1706645.
- [5] D. Andrew Austin and T. L. Hungerford, "The market structure of the health insurance industry," *Heal. Insur. Ind. Mark. Struct. Cover. Cost Issues*, pp. 1–79, 2011.
- [6] A. Ho, "Health Insurance," *Encycl. Glob. Bioeth.*, no. June 2015, 2020, doi: 10.1007/978-3-319-05544-2.
- [7] D. T. Declarations, "How to Read an Insurance Policy," vol. 20036, no. 202, 2004, [Online]. Available: <https://nonprofitrisk.org/>.
- [8] L. Macedo, "The Role of the Underwriter in Insurance," *Prim. Ser. Insur.*, vol. 1, no. 8, pp. 13–29, 2009.
- [9] O. Espinosa and A. Zarruk, "The importance of actuarial management in insurance business decision-making in the twenty-first century," *Br. Actuar. J.*, vol. 26, no. 1, pp. 1–16, 2021, doi: 10.1017/S1357321721000155.
- [10] S. Dyson, B. Hardy, and B. Leung, "The role of the actuary in healthcare: where are we, and where are we going?," *2003 Bienn. Conv. - Shap. Futur. a world Uncertain.*, no. May, pp. 1–20, 2003.
- [11] J. J. Sun, "Identification and Prediction of Factors Impact America Health Insurance Premium," 2020.
- [12] P. Embrechts and M. V. Wüthrich, "Recent Challenges in Actuarial Science," *Annu. Rev. Stat. Its Appl.*, vol. 9, no. 1, pp. 119–140, 2022, doi: 10.1146/annurev-statistics-040120-030244.
- [13] X. Landes, "How Fair Is Actuarial Fairness?," *J. Bus. Ethics*, vol. 128, no. 3, pp. 519–533, 2015, doi: 10.1007/s10551-014-2120-0.
- [14] C. Berryman, C. J. Ferguson, and C. Negy, "Social Media Use and Mental Health among Young Adults," *Psychiatr. Q.*, vol. 89, no. 2, pp. 307–314, 2018, doi: 10.1007/s11126-017-9535-6.
- [15] N. Ittiprasert, "The application and efficiency of digital marketing strategies of life insurance market in Thailand," 2021, vol. 37, pp. 76–112.
- [16] C. Lee Ventola, "Social media and health care professionals: Benefits, risks, and best practices," *P T*, vol. 39, no. 7, pp. 491–500, 2014.
- [17] R. Skaik and Di. Inkpen, "Using Social Media for Mental Health Surveillance: A Review," *ACM Comput. Surv.*, vol. 53, no. 6, 2021, doi: 10.1145/3422824.
- [18] G. T. and M. A. Ahad, "Effects of Social Media on Social, Mental, and Physical Health Traits of Youngsters," *Springer Nat. Singapore Pte Ltd*, vol. 11, pp. 685–695, 2019, doi: 10.1007/978-981-10-8055-5_61.
- [19] J. P. L. Kenneth C. Laudon, *MANAGEMENT INFORMATION SYSTEMS*. New York: Pearson Education, Inc, 2020.
- [20] T. Nguyen *et al.*, "Prediction of Population Health Indices from Social Media using Kernel-based Textual and Temporal Features," 2017, pp. 99–107, doi: 10.1145/3041021.3054136.
- [21] R. T. & P. Wolffl, "Predicting future mental illness from social media: A big-data approach," *Behav. Res. Methods*, vol. 15, pp. 1586–1600, 2019, doi: 10.3758/s13428-019-01235-z.
- [22] O. Gruebner, M. Sykora, S. R. Lowe, K. Shankardass, S. Galea, and S. V. Subramanian, "Big data opportunities for social behavioral and mental health research," *Social Science and Medicine*, vol. 189, pp. 167–169, 2017, doi: 10.1016/j.socscimed.2017.07.018.
- [23] C. Srisawatsakul and W. Boontarig, "An assessment of privacy concerns on personal health information: Thailand case study," *Curr. Appl. Sci. Technol.*, vol. 21, no. 4, pp. 774–787, 2021, doi: 10.14456/cast.2021.62.
- [24] "EU general data protection regulation," [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>.
- [25] M. Shuaib, S. Alam, M. Shabbir Alam, and M. Shahnawaz Nasir, "Compliance with HIPAA and GDPR in blockchain-based electronic health record," *Mater. Today Proc.*, no. xxxx, 2021, doi: 10.1016/j.matpr.2021.03.059.
- [26] D. M. De Simone, "When Is Accessing Medical Records a HIPAA Breach?," *J. Nurs. Regul.*, vol. 10, no. 3, pp. 34–36, 2019, doi: 10.1016/S2155-8256(19)30146-2.
- [27] U. S. D. of H. & H. S. Office for Civil Rights, "OCR PRIVACY BRIEF SUMMARY OF THE HIPAA PRIVACY RULE HIPAA Compliance Assistance," *Summ. HIPAA Priv. Rule*, p. 23, 2003, [Online]. Available:

www.hhs.gov/sites/default/files/privacysummary.pdf.

- [28] J. Kaewkungwal, P. Adams, J. Sattabongkot, R. K. Lie, and D. Wendler, "Issues and challenges associated with data-sharing in LMICs: Perspectives of researchers in Thailand," *Am. J. Trop. Med. Hyg.*, vol. 103, no. 1, pp. 528–536, 2020, doi: 10.4269/ajtmh.19-0651.
- [29] Kingdom of Thailand, "Thailand Personal Data Protection Act, B.E. 2562, 2019," no. 136, pp. 1–35, 2019, [Online]. Available: <https://thainetizen.org/wp-content/uploads/2019/11/thailand-personal-data-protection-act-2019-en.pdf>.
- [30] M. Conway and D. O'Connor, "Social media, big data, and mental health: Current advances and ethical implications," *Curr. Opin. Psychol.*, vol. 9, pp. 77–82, 2016, doi: 10.1016/j.copsyc.2016.01.004.
- [31] B. Custers, S. van der Hof, B. Schermer, S. Appleby-Arnold, and N. Brockdorff, "Informed Consent in Social Media Use – The Gap between User Expectations and EU Personal Data Protection Law," *SCRIPTed*, vol. 10, no. 4, pp. 435–457, 2013, doi: 10.2966/scrip.100413.435.
- [32] S. Lomborg and A. Bechmann, "Using APIs for Data Collection on Social Media," *Inf. Soc.*, vol. 30, no. 4, pp. 256–265, 2014, doi: 10.1080/01972243.2014.915276.
- [33] A. Le Glaz *et al.*, "Machine learning and natural language processing in mental health: Systematic review," *J. Med. Internet Res.*, vol. 23, no. 5, 2021, doi: 10.2196/15708.
- [34] O. Baclic, M. Tunis, K. Young, C. Doan, and H. Swerdfeger, "Challenges and opportunities for public health made possible by advances in natural language processing," *Canada Commun. Dis. Rep.*, vol. 46, no. 6, pp. 161–168, 2020, doi: 10.14745/ccdr.v46i06a02.
- [35] B. L. Cook, A. M. Progovac, P. Chen, B. Mullin, S. Hou, and E. Baca-garcia, "Novel Use of Natural Language Processing (NLP) to Predict Suicidal Ideation and Psychiatric Symptoms in a Text-Based Mental Health Intervention in Madrid," vol. 2016, 2016.
- [36] C. Y. Ku, P. C. Sung, and W. H. Hsieh, "Policy satisfaction for separation of dispensing from medical practices in Taiwan: Success of the prescription-release information system," *Telemat. Informatics*, vol. 31, no. 2, pp. 334–343, 2014, doi: 10.1016/j.tele.2013.09.004.
- [37] V. K. Shrotryia and U. Dhanda, "Content Validity of Assessment Instrument for Employee Engagement," *SAGE Open*, vol. 9, no. 1, 2019, doi: 10.1177/2158244018821751.
- [38] E. Almanasreh, R. Moles, and T. F. Chen, "Evaluation of methods used for estimating content validity," *Res. Soc. Adm. Pharm.*, vol. 15, no. 2, pp. 214–221, 2019, doi: 10.1016/j.sapharm.2018.03.066.
- [39] OECD.org, "Evaluation Criteria," 2018. <https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm> (accessed Feb. 15, 2022).



Nattarika Poolperm was born in Ubonratchathani, Thailand in 1989. She received the Bachelor of Arts in Business English from Khon Kaen University, Khon Kaen, Thailand in 2012. She is currently a master's degree student in Information Technology Management (ITM) Mahidol University. Her area of research interests are global health and data analysis.



Theeraya Mayakul received a Pharm. D. degree in pharmaceutical care from Prince of Songkla University, Thailand in 2009. She was a clinical pharmacist at Siriraj hospital from 2009 to 2016. She received an M.Sc. degree in Information Technology Management (ITM) from Mahidol University in 2015. She is currently Ph.D in ITM. Her areas of research interests are medical informatics, pharmaceutical application, enterprise architecture, and information management.



Supaporn Kiattisin received the B.Eng. degree in applied computer engineering from the Chiang Mai University, Chiang Mai, Thailand, in 1995, the M.Eng. degree in electrical engineering, and the Ph.D. degree in electrical and computer engineering from King Mongkut's University of Technology Thonburi, Bangkok, Thailand, in 1998, and 2006. She is currently the program director of Information Technology Management, Faculty of Engineering, Mahidol University, Thailand. Her research interests include computer vision, image processing, robot vision, signal processing, pattern recognition, artificial intelligence, IoT, IT management, digital technologies, big data, and enterprise architecture with TOGAF 9 certified. She is a member of IEICE and TESA. She served as a Head of IEEE Thailand Chapter in Bio-Medical Engineering. She also served as the Chairman of the TimesSOC Transaction Thailand



Somkiat Wattanasirichaigoon received the B.Sc. in Biomedical Science in 1981, B.M. Medicine in 1983, and B.Sc. in Clinical Medical Sciences in 1986 from Mahidol University, Thailand. Besides, he graduated his B.P.A in Public Administration in 1992 from Sukhothai Thammathirat Open University. Then, in 1988 he completed in Surgery, Diploma of the Thai Subboard of Medicine from the Medical Council of Thailand. He is currently professor MD of Information Technology Management, Faculty of Engineering, Mahidol University, Thailand. His areas of research interest are health informatics, healthcare information technology management and business planning and strategy.