International Journal of Electrical and Computer Engineering (IJECE)
Vol. 12, No. 6, December 2022, pp. 6387~6396
ISSN: 2088-8708, DOI: 10.11591/ijece.v12i6.pp6387-6396

Cognitive level classification on information communication technology skills for blog

Chalothon Chootong, Jirawan Charoensuk
Department of Computer Science and Technology, Faculty of Science at Sriracha, Kasetsart University, Chonburi, Thailand

Article Info

Article history:

Received Sep 28, 2021
Revised Jun 14, 2022
Accepted Jul 10, 2022

Keywords:

Cognitive level
Deep neural network
Information communication technology skills
Medium blog
Skill classification

ABSTRACT

Learners can study and update their knowledge continually due to the rapid growth of online content. The Medium blog is a well-known open platform that encourages authors who want to share their experiences to publish content on various topics in multiple languages. Meanwhile, readers can query interesting content by searching for a related topic. However, finding suitable content is still challenging for learners, especially information communication technology (ICT) content in Thai, and needs to be classified into beginner, intermediate, and advanced cognitive levels. Moreover, ICT blog content is usually a mix of Thai language and technical terms in English. To overcome the challenge of content classification, a deep neural network (DNN) classification model was constructed to classify the ICT content from the Medium blog into three levels based on cognition. We examined and compared the classification results with strong baseline models, including logistic regression, multinomial naïve bayes, support vector machine (SVM), and multilayer perceptron (MLP). The experimental results indicate that the proposed DNN model attained the highest accuracy (0.878), precision (0.882), recall (0.878), and F1-score (0.875).

Corresponding Author:

Jirawan Charoensuk
Department of Computer Science and Technology, Faculty of Science at Sriracha, Kasetsart University
Sukhumvit Road, Thungsukla Subdistrict, Sriracha District, Chonburi 20230, Thailand
Email: jirawan.charo@ku.th

1. INTRODUCTION

The popularity and utilization of online learning have rapidly increased. A learner can study from any location via many learning resources (e.g., video, slides, blogs, graphics, and animation) using technologies through the internet. Presently, online learning is an essential resource for learners, especially information and communication technology (ICT) skills that are regularly changing. ICT skills can help people to work more efficiently and facilitate jobs by carrying out tasks such as recording the time in and out of work, sharing documents on the cloud, video recording meetings, and so on. It has become indispensable, especially for students who want to learn and develop their skills and workers who want to upskill and reskill. However, finding appropriate content for the learner's knowledge level is challenging, and matching the learner with suitable content is crucial.

Bloom's taxonomy is a model used to classify educational learning by using the six cognitive hierarchy levels: remembering, understanding, applying, analyzing, evaluating, and creating, which are used to organize academic learning and recommending appropriate content to learners with different knowledge levels. Many researchers have been working on content recommendations to address the problem of recommending relevant and suitable content for learners. In study [1], learner patterns were used to construct a content recommendation system based on fuzzy logic. Knowledge units can be used as a guideline to build models for recommending suitable content [2], although the knowledge domain limits their usefulness. Recently, machine learning algorithms have been applied to classify e-learning content into various cognitive levels [3], [4] and for text analysis [5]. Most of the current work is focused on English learning content. Although [3], [6] studied the cognitive levels of the Thai language by using Bloom's taxonomy, it cannot be directly applied to ICT skill cognitive levels or the Thai paradigm. Moreover, existing learning content classification systems have been built based on historical user data, which makes the system limited. Using a fixed list of words to classify the content level is not proper for the ICT skill domain since it constantly changes. To address this problem, we identified the content level by observing content combinations. A deep learning model was trained to understand the content relationship and classify the blog content satisfactorily.

The aim of this study is to determine appropriate ICT knowledge for learners at the cognitive level. We introduce a framework of ICT skills classification in the Thai language by using the Medium blog as the data source (2,213 in total). Our framework is organized into three main processes: skill collection, skill tagging, and skill level classification. We created a function to scrape the blog content with keyword sets for the skill collection process, as shown in Algorithm 1. Furthermore, the skills tagging system was designed as a tool for preparing and rechecking the dataset. Finally, we analyzed the experimental results based on five machine and deep learning classifier models to assess their precision, recall, F-measure, and accuracy. The results clearly show that the deep neural network (DNN) yielded the best performance for cognitive level classification of ICT skills in Thai language blogs. Our contributions are summarized: i) we introduce the ICT blog content dataset populated via a crawler function and labeled by experts, ii) we proposed a framework to classify the cognitive level of the ICT blog content into beginner, intermediate, and advanced knowledge levels, and iii) we carried out experiments with machine and deep learning techniques and report the comparison results.

2. LITERATURE REVIEW

2.1. Learning blogs

The ever-increasing improvement in technology means that learners can access content anywhere at any time. Meanwhile, content creators can create subject matter and share knowledge via various channels such as websites, YouTube, Facebook, and Blogs. Blogs generally include a comments feature that allows readers to enter discussions with the blog's author, and a reader can show their appreciation for the author by rating or giving a score. Meanwhile, readers can post and share blogs on social media or save them to read later. In the past, blogs have been used to create effective learning environments that allow users and learners to have deeper thought processes, enjoy the interaction, and progress well. Indeed, using blogs in the classroom can support reflective learning [7]. In Sidek and Yunus [8] collected data from observing students' journal entries and comments and demonstrated learning improvement of students by using blogs as learning journals. Furthermore, blogs can be used as a learning tool for enhancing skills [9]. Currently, blogs are not only used in the classroom but also on public platforms. The Medium blog is an open platform where readers can find many kinds of content and where experts can share their knowledge or experience on any topic. The content is written in various languages, including Thai, and covers multiple topics such as arts and entertainment, culture, health, industry, politics, programming, and technology [10]. It can be the principal resource that readers use as the material to study or update their skills, primarily in the ICT domain. The blog has more than 100 million readers worldwide [11].

2.2. Cognitive levels

Bloom's taxonomy is a set of hierarchical models (cognitive, affective, and psychomotor) used to organize educational learning objectives into levels of difficulty and specification [12]. Later, Bloom's taxonomy was revised by changing level names and ordering the cognitive hierarchy level. Thus, the new Bloom's taxonomy comprises remembering, understanding, analyzing, evaluating, and creating [13]. Cognitive skills are essential for improving learning performance [14]. The cognitive level of a learner is also related to the level of learning content. Many authors have analyzed the cognition of Thai students based on Bloom's taxonomy. However, classifying the content level is still challenging, especially the content written in Thai. Aninditya *et al.* [3] proposed categorization of exam questions based on the cognitive level divided into lower and high orders by using Bloom's taxonomy. The lower order contains three levels from the lower levels of Bloom's taxonomy: remembering, understanding, and analyzing, with the remaining levels of Bloom's taxonomy (evaluating and creating) being defined as high order. They used midterm and final exam questions from the Department of Information Systems as the dataset, categorized the learning objectives, and then provided exam questions for evaluating students. In Thomas and Chandra [4] purposed the framework which recommends appropriate learning content to a user of an e-learning system. This framework also focused on the cognitive level and separated learning content into three different levels:

beginner level, which consists of remembering and understanding; intermediate level, which consists of applying and analyzing; and advanced level, which consists of evaluating and creating. Anekboon [6] classified questions into six levels of cognitive taxonomy in Thai. They solved ambiguous Thai words and phrases and focused on feature selection techniques specified by relevance. Moreover, the dataset comprised general knowledge exams from websites.

2.3. Machine and deep learning for text classification

Extracting insightful information from text documents to elucidate its main idea and to classify the type of content can be challenging and time-consuming. Machine learning methods have been utilized to learn the document's content and classify it into relevant categories. Shah *et al.* [15] compared the performances of three machine learning models: logistic regression, random forest (RF), and k-nearest neighbors (KNN) for text classification into various categories; business, entertainment, politics, sports, and technology. They found that the logistic regression classifier with the term frequency-inverse document frequency (TF-IDF) vectorizer feature achieved the highest accuracy on a particular dataset. Naïve Bayes has been used in both industry and academia for a long time for information retrieval and text classification [16], [17]. Arreerard and Senivongse [18] studied the efficiency of two machine learning methods: support vector machine (SVM) and naïve Bayes for classifying defamatory text in Thai with several combinations of approaches. Their results show that SVM performed better than naïve Bayes.

In recent years, deep learning techniques have been applied for text classification, including sentiment analysis, news categorization, question, and answer, and so on [19]. Neural networks can be used to automatically learn unstructured and complex language patterns, especially the Thai language structure. Piyaphakdeesakun *et al.* [20] examined a deep learning approach for analyzing online documents in Thai. They focused on two problems with the Thai language: ambiguous word manipulation and automatic sentiment classification. They found that bi-directional gated recurrent unit (GRU) with an attention mechanism produced the best performance for sentiment classification. Similarly, Thai sentiment analysis using hybrid deep learning models was studied by Pasupa and Ayutthaya [21].

3. RESEARCH METHOD

Herein, we introduce content classification for ICT skills from the Medium blog in Thai. Figure 1 presents the proposed framework. It is composed of three main processes: skill collection, skill tagging, and skill level classification.

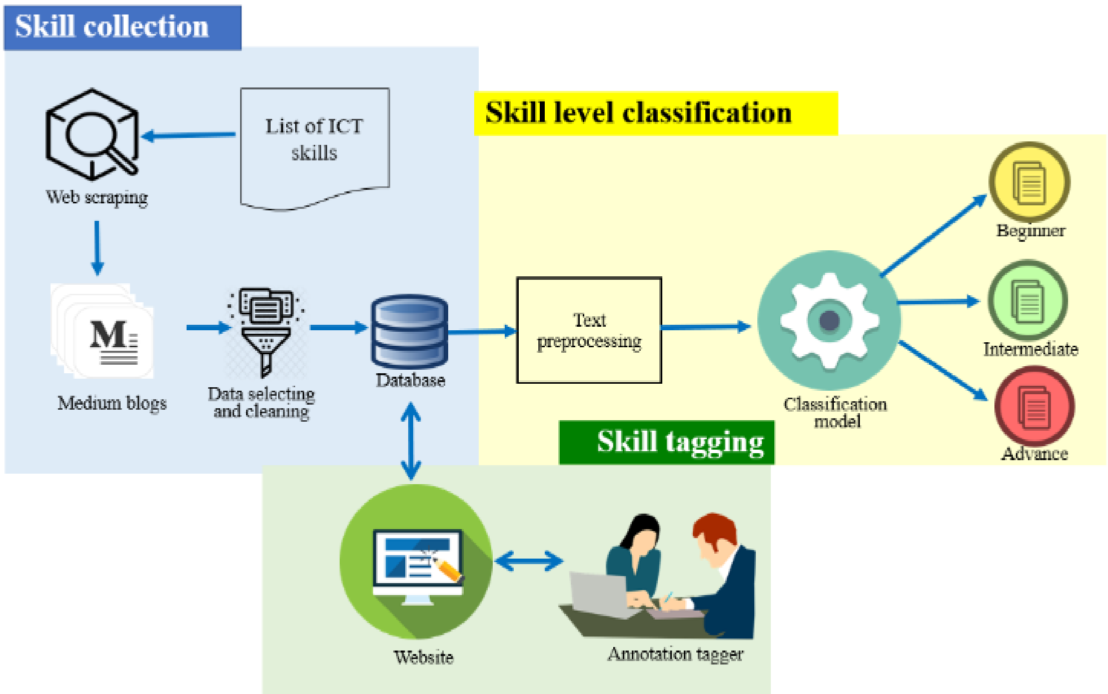

Figure 1. The framework for content classification of ICT skills

3.1. Skills collection

To prepare the dataset, we specifically focused on ICT skills on the Medium blog in Thai. From the observations, Thai authors generally set the title of the Medium blog by using ICT terms such as programming language, platform, tools, and software in both English and Thai. There are eight sections of a Medium blog: title, author's name, publication date, text content, first blockquote of the blog, list of tags, number of claps, and readers' reviews

In the data collection process, the sets of keywords for the web crawlers were constructed by using algorithm 1 as Figure 2. First, the list of keywords (*KWs*) was set to an empty list, TH_{terms} is a list of skills in Thai, and EN_{terms} is a list of skills in English. To extend the set of keywords, second, we combined EN_{terms} with 44 Thai alphabet letters as $alpha_i$. The combination results were stored in $KW_{Alphabet}$. Besides, we constructed a list of 45 specific terms ($Special_{terms}$), such as "ติดตั้ง" (install), "เรียน" (study), and "ตัวอย่าง" (example), with each word represented by s_i, to enhance the crawler keywords; these results were stored in $KW_{Special}$. Finally, we merged all TH_{terms}, $KW_{Alphabet}$, and $KW_{Special}$ to generate the list of *KWs* for the crawler process.

Algorithm 1: Crawler's keywords preparation

Initialization $KWs \leftarrow \emptyset, KW_{Alphabet} \leftarrow \emptyset, KW_{Special} \leftarrow \emptyset$
$TH_{terms} = set\ of\ ICT\ terms\ in\ Thai,$
$EN_{terms} = set\ of\ ICT\ terms\ in\ English,$
$Special_{terms} = set\ of\ specific\ terms$
For $term \in EN_{terms}$ **do**
    For $alpha_i \in Thai\ Alphabets$ **do**
        $KW_{Alphabet} \leftarrow KW_{Alphabet} \cup Concat(alpha_i, term)$
    End
    For $s_i \in Special_{terms}$ **do**
        $KW_{Special} \leftarrow KW_{Special} \cup Concat(s_i, term)$
    End
$KWs = TH_{terms} \cup KW_{Alphabet} \cup KW_{Special}$
Return KWs

Figure 2. Pseudocode of skills collection

3.2. Skills tagging

We defined cognitive levels into three classes as used in [4] and defined the ICT skill levels. A beginner level blog is one in which the author includes basic knowledge about an ICT skill that the reader can understand without any previous knowledge. For the intermediate level, the blog content covers the relationship between two skills with some background knowledge. For the advanced level, the blog comprises mixed content that explains how to apply ICT skills in a case study or reports on a workshop example. To train the model, we prepared the ground truth by using the outputs from the skill tagging process. Accordingly, annotation comprised two levels: related labeling consisting of three types (undefined, yes, and no) and content level labeling consisting of three types (beginner, intermediate, and advance). Blogs were annotated by using two groups of annotation taggers familiar with ICT skills. The first group comprised students in computer science and information technology programs. The secondary group is an expert in ICT skills. Labels with most votes were selected. Related labeling was used to manage the level of ICT skill relevance.

When the first group of taggers could not decide on the level of the blog content, they labeled it as "Undefined". This data was then sent to the experts for defining the label for the content level. When the "Yes" label was assigned to a blog, the taggers could specify a label for the content level whereas when "No" was assigned, and the blog was removed. Content level labeling set the level of the ICT skill content as beginner, intermediate, or advanced. All steps for skills tagging are illustrated in Figure 3. The annotation taggers defined the skill level of each blog, which were then stored in the database to be used for further steps. Examples of blog content labeling are summarized in Table 1.

3.3. Skill level classification

We studied and experimented with various techniques to construct the level classification model. In this paper, we select logistic regression, multinomial naïve Bayes, SVM, multilayer perceptron (MLP) and DNN to construct this model. We briefly describe machine learning algorithms in this section.

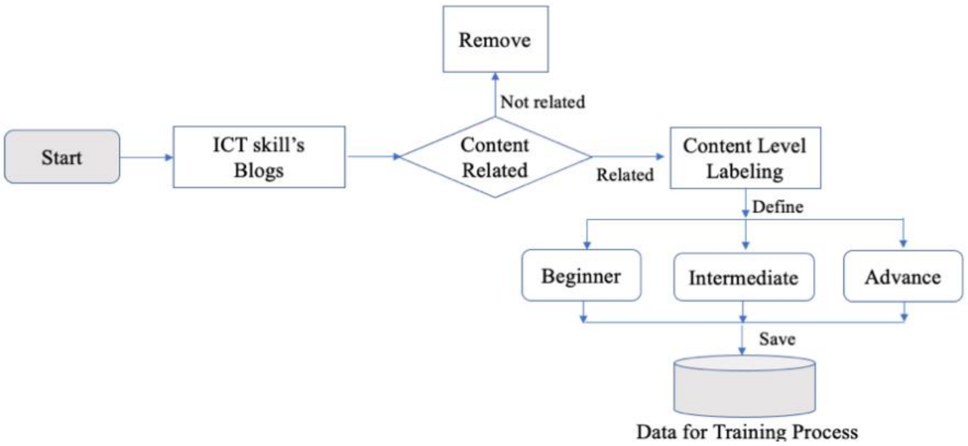

Figure 3. The steps for skills tagging

Table 1. Blog content results and examples of skill labeling

Skill Label	Number of Blogs	Example
Beginner	975	รีวิว SIPA Programming เป็น Channel YouTube สอนเขียนโปรแกรม Python เบื้องต้น ฟรี ["A review of SIPA Programming, which is a YouTube channel teaching an introduction to Python programming for free."]
Intermediate	502	พัฒนาเว็บแอพพลิเคชันด้วยภาษา Python|Python Flask ["Develop web applications with Python | Python Flask."]
Advance	736	ใช้ Python ดึงข้อมูล API ของ Coronavirus (COVID-19) ["Use Python to retrieve data from the Coronavirus (COVID-19) API."]
Total	2,213	

3.3.1. Logistic regression

Logistic regression is a type of classification algorithm commonly used for the classification baseline method. It is a predictive analysis algorithm using probability via a Sigmoid function as a cost function to train with the data. The hypothesis representation of logistic regression slightly modified from linear regression that gives values between 0 and 1 that can be calculated by (1):

$$h\theta(X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}} \quad (1)$$

where *X* is the feature vector of content from the Medium blog produced via the TF-IDF technique.

3.3.2. Multinomial naïve Bayes

For multinomial naïve Bayes, the Medium blog content was represented by a feature vector where the elements of the features are expressed as the frequency with which term *i* occurs in the content. Thus, the probability of the content of the blog (*X*) belonging to class *j* (C_j)is estimated as (2):

$$p(X|C_j) = \prod_{i=1}^{n} p(t_i|C_j). \quad (2)$$

$p(t_i|C_j)$ is calculated by (3),

$$p(t_i|C_j) = \frac{N_{t_i,C_j}+\alpha}{N_{C_j}+\alpha n} \quad (3)$$

where N_{t_i,C_j} is the number of t_i occurs in the training set of class C_j, N_{C_j} is the total number of all terms in class $C_j$and $\alpha$is a smoothing parameter.

3.3.3. Support vector machine

SVM, a machine learning algorithm widely used in text classification, comprises supervised learning models described by using decision hyperplanes [22]. The weight vector ($\overrightarrow{W}$) and bias term *b* provide the best decision hyperplane by solving the optimization problem. The data ($\vec{X}_i$) can be classified by using (4).

$$f(\vec{x}_i) = sign(\overline{w}^T \vec{x}_i + b). \quad (4)$$

3.3.4. Multilayer perceptron

MLP is a primary neural network. In this study, we built the labeling model based on MLP from [23]. The model parameters were configured with *hidden_layer_sizes*=128, *max_iter* for the number of epochs=50, activation='ReLU', and solver='adam' for weight optimization. Subsequently, we compared the model efficiency with other techniques. The model optimizes the log-loss function by using stochastic gradient descent by updating the weights as (5):

$$w = w - (lr * w_{grad}) \quad (5)$$

where w is the weight (which is updated), lr is the learning rate (set to 0.001 in our study), and w_{grad} is the gradient of w with applied loss.

3.3.5. Deep neural network

In this section, we explain the overall process of the skills level model shown in Figure 4. The proposed model consists of an embedded layer, which takes token vectors as the input for the network, and four dense layers, the last one being for level content prediction. Tokens were generated by using the unigram and bigram of a word, after which we computed the token frequency and normalized it by using TF-IDF weight. To solve the problem of the optimal local solution, we added three dense layers to the network. In addition, we used rectified linear unit (ReLU) instead of sigmoid to solve the problem of gradient disappearance. Moreover, we used Adam for optimization to help to find appropriate weights during the training of the model.

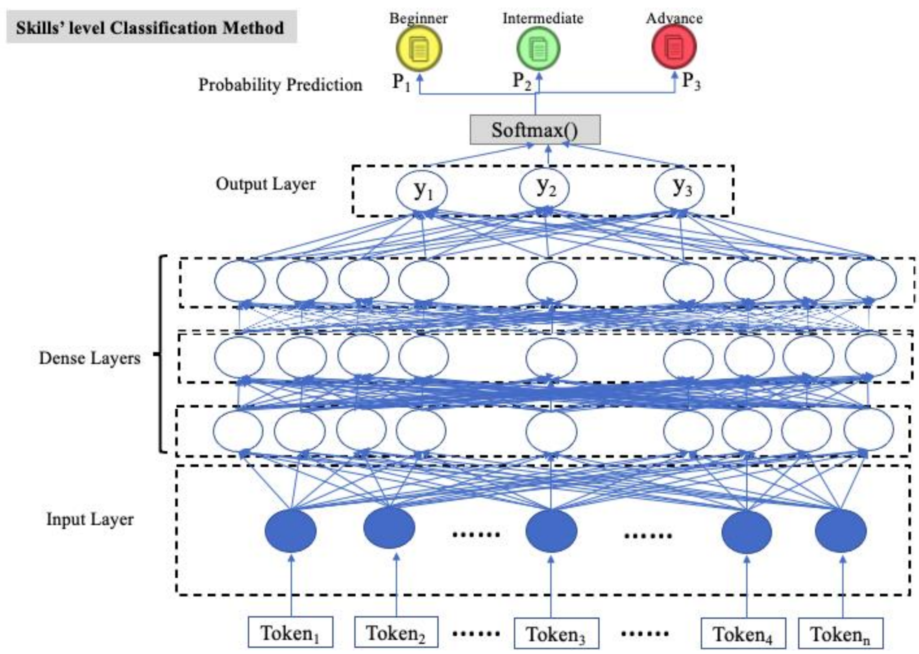

Figure 4. The skill level classification method

The configuration of parameters in the network are: i) the input layer size was the max number of tokens represented by n; ii) the hidden layer size was 512, and we adopted ReLU as the activation function; iii) dropout was set as 0.5 to reduce the risk of overfitting; iv) the number of nodes was set as 3 for the output layer of the three levels (Beginner=0, Intermediate=1, Advanced=2) to label the blog content. The softmax function was applied as an activation function for the output layer; and v) sparse categorical cross entropy was used as the loss function compute by (6),

$$Loss = -\sum_{i=1}^{output\_size} y_i \log \hat{y}_i \tag{6}$$

where y_i is the true class index and $\hat{y}_i$ is the predicted class index. One benefit of applying sparse categorical cross entropy is it saves time in memory as well as computation.

3.4. Implementation

3.4.1. Data selection and cleaning

The data cleaning process includes four functions: replacing empty sections with null, removing adverts, removing content without text, and removing repeated blogs. The cleaned data were then saved to the database for the skill tagging process. The retrieved blogs are related with three categories ICT skills include programming languages, platforms, tools, and software. After cleaning process, our dataset contains 2,213 blogs that ready to train model.

- Replacing empty sections with null: because a Medium blog does not force the author to write content in all sections, some sections such as the publication date, list of tags, numbers of claps are empty, and these are replaced with null.
- Removing non-letter: blog authors usually explain their content by using text, images, and videos together. Our focus was on text content, and so blog non-letters were removed.
- Removing repeated blogs: keyword searches during the data scraping produced repeated results for the same blog with the same skills. The repeated blogs were checked and removed by this function.

3.4.2. The pre-trained process

Text preprocessing was used to select blog content and transform texts into feature vectors for generating the classification model. Although blogs usually comprise text, images, video clips, hyperlinks, we were only interested in the text content. For text preprocessing, word tokenization, lowercase transformation, and stop-word removal were conducted. Table 2 reports the sequence of the text preprocessing steps with some Medium blog content as examples.

a. Word tokenization

Symbols to identify the boundaries of a sentence and spaces to identify the boundaries of a word used in English are not used in Thai. The Medium blog text content comprises unstructured text and free-form styles. When creating the content, the authors usually use ICT skills in both Thai and English, which may cause segmentation problems. Word segmentation was achieved by using a technique based on combining a dictionary based on the maximum matching algorithm [24] and name entity recognition (NER) [25]. The maximum matching algorithm searches for sequences of characters and matches them with words in the dictionary. Subsequently, the list of words is sent to NER and combined with a list of tuples via the conditional random fields algorithm.

b. Lowercase transformation

The names of programming languages, platforms, tools, and software are typically in English. We transformed all English words to lowercase. Then, it constructs the feature vector generation process more efficient.

c. Stop-word removal

Stop-words commonly occur in texts and are not required for the skills classification process, and so the role of this step is to filter them out of the text. There are 113 of them: adverbs (e.g., “เฉพาะ”, “ขึ้น”), auxiliary verbs (e.g., “จะ”, “ต้อง”), conjunctions (e.g., “และ”, “หรือ”), determiners (e.g., “นี่”, “นั้น”), prepositions (e.g., “โดย”, “กับ”), particles (e.g., “ค่ะ”, “เป็นต้น”), and punctuation (e.g., “และ”, “หรือ”).

Table 2. The output from the text preprocessing steps

<table>
<tr><th>Process</th><th>Result</th><th>Remark</th></tr>
<tr><td>Input</td><td>รีวิว_SIPA_Programming_เป็น_Channel_YouTube_สอนเขียนโปรแกรม_Python_เบื้องต้น_ฟรี
[“A review of SIPA Programming, which is a YouTube channel teaching an introduction to Python programming for free.”]</td><td>Underscores (_) represent spaces.</td></tr>
<tr><td>Word tokenization</td><td>รีวิว|SIPA Programming|เป็น|Channel_YouTube|สอนเขียนโปรแกรม|Python|เบื้องต้น| ฟรี|</td><td>8 words and vertical bar (|) represents a segmented sign</td></tr>
<tr><td>Lowercase transformation</td><td>รีวิว|sipa programming|เป็น|channel_youtube|สอนเขียนโปรแกรม|python|เบื้องต้น|ฟรี|</td><td>set the English content to lower case.</td></tr>
<tr><td>Stop-word removal</td><td>รีวิว|sipa programming|channel_youtube|สอนเขียนโปรแกรม|python |เบื้องต้น| ฟรี|</td><td>remove 1 word; เป็น (is).</td></tr>
</table>

4. RESULTS AND DISCUSSION

Feature extraction generates a feature vector by using unigrams and bigrams, while TF-IDF is used to assign weights to the items in the vector. We built five classification methods: logistic regression (LR), multinomial naïve Bayes, SVM, MLP, and DNN. For training the model, the data were randomly split into 80% for training and 20% for testing. We examined the classification results of the blog content based on the five models. Scoring methods such as precision, recall, F1-score, and accuracy were applied to measure the framework's performance and effectiveness.

The experimental results of skills level classification are listed in Table 3. We first discuss the results for the first three models: LR, multinomial naïve Bayes, and SVM, which are well-known and often-used machine learning algorithms. Of these, SVM performed the best with the highest precision, recall, F1-score, and accuracy scores of 0.822, 0.823, 0.812, and 0.822, respectively. When comparing the other two methods, we found that DNN outperformed MLP and reached the highest precision, recall, F1-score, and accuracy scores of 0.882, 0.878, 0.875, and 0.878, respectively.

Table 3. Experimental results of skill level classification

Models	Measurement Metric (%)			
	Precision	Recall	F1-score	Accuracy
Linear Regression	0.762	0.771	0.763	0.771
Multinomial Naïve Bayes	0.754	0.731	0.743	0.732
Support Vector Machine	0.822	0.823	0.812	0.822
Multilayer Perceptron	0.801	0.803	0.800	0.803
Deep Neural Network	0.882	0.878	0.875	0.878

Figure 5 shows the confusion matrix for the DNN model results. The confusion matrix scores clearly show that the model could correctly classify blog content in Thai even at the advanced cognitive level. However, there were some incorrect results, and there are two main reasons that might have caused incorrect predictions. Some intermediate level blog content contained terms that usually occur in beginner level content, such as *("พื้นฐาน," "basic"), ("เริ่ม," "start"), ("สิ่งแรก," "the first thing")*. Blog content should be identified as intermediate level when it contains more than two techniques or is part of the next step of learning an ICT skill, including sentences such as "*เริ่มต้นสร้างฐานข้อมูลของเรากันเถอะ (Let' s start to build the database),*" "*ก่อนอื่นต้องเปิดใช้งาน PHP และ MySQL Database (First, we have to set the configurations for PHP and the MySQL database)*" [26]. The second reason is blog authors usually include the content of the previous episode in the first paragraph of the new one as a review, which can cause ambiguity in the Skill labeling between the beginner and intermediate levels.

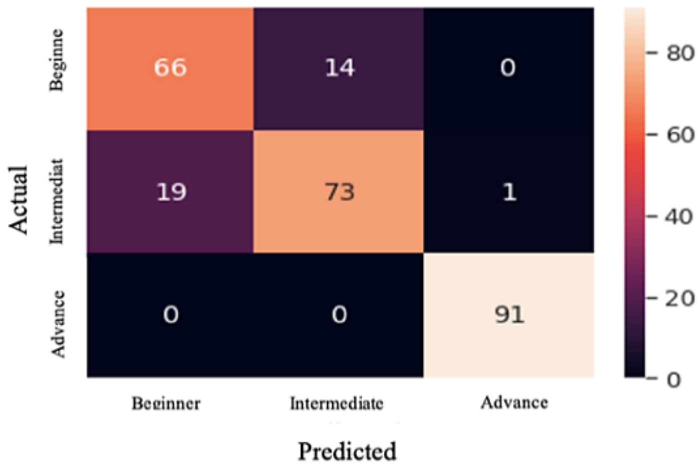

Figure 5. The confusion matrix for the deep neural network model

A web-based graphical user interface (GUI) was created to use in the skills level classification of the ICT as shows in Figure 6. A user inputs two data items: the title and content of the blog into the web page and clicks the predict button. Our trained model returns output that is content level at the top of the webpage.

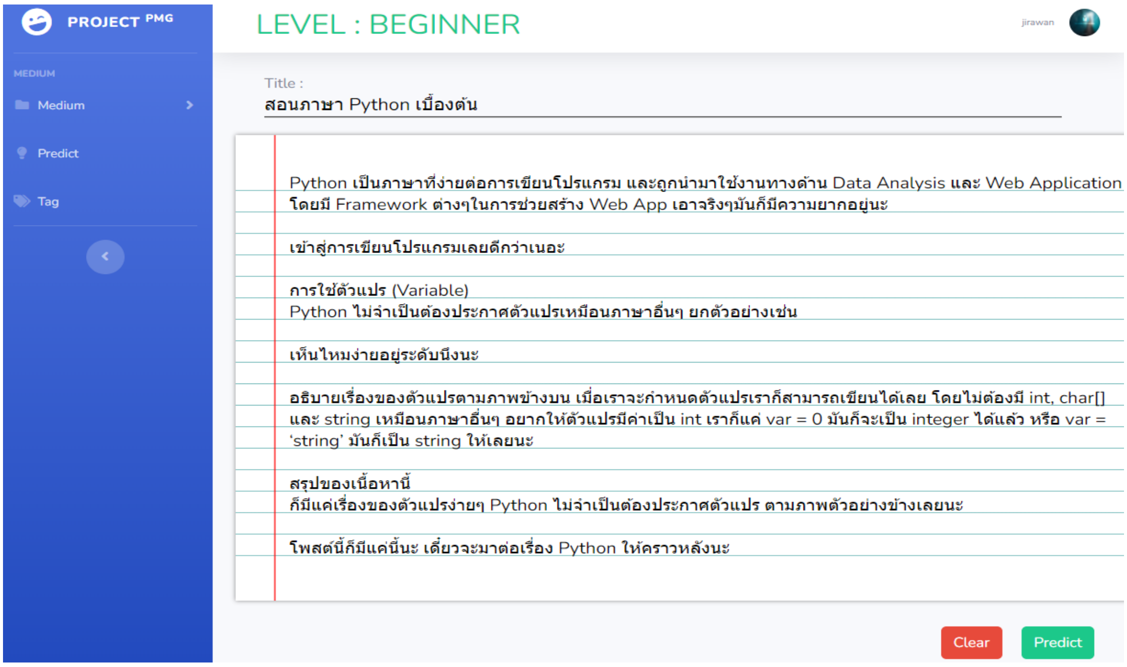

Figure 6. An example of skill level prediction

5. CONCLUSION

Herein, we proposed a framework for Thai language content classification of ICT skills from the Medium blog. The framework consists of the three main processes: skill collection, skill tagging, and skill level classification. First, skill collection was used to collect four types of ICT skills: programming languages, platforms, and tools, and software from the Medium blog. We developed an algorithm to crawl and collect blogs in both Thai and English because Thai authors usually set blog titles using both languages for the ICT skill terms. The second process, Skill tagging was developed to identify and store data for the next process. We used the cognitive skills from Bloom's taxonomy and classified the skill labels into three types: beginner, intermediate, and advanced. Two groups of annotation taggers, students and experts in computer science, tagged the skill levels for each blog. Finally, the skill level classification model based on DNN was used to identify correctly labeled ICT skills blogs. We evaluated its performance by comparing its classification ability with those of five models: linear regression, multinomial naïve Bayes, and SVM, MLP, and deep neural network. The experimental results show that DNN performed the best with 0.882, 0.878, 0.875, and 0.878 scores for precision, recall, F1-score, and accuracy, respectively, and yielded the best performance for identifying the advanced labels.

ACKNOWLEDGEMENTS

This research was also supported by the Faculty of Science at Sriracha, Kasetsart University.

REFERENCES

[1] S. Pariserum Perumal, G. Sannasi, and K. Arputharaj, "An intelligent fuzzy rule-based e-learning recommendation system for dynamic user interests," *The Journal of Supercomputing*, vol. 75, no. 8, pp. 5145–5160, Aug. 2019, doi: 10.1007/s11227-019-02791-z.

[2] F. Nafa, J. I. Khan, and S. Othman, "Extending cognitive skill classification of common verbs in the domain of computer science for algorithms knowledge units," in *Proceedings of the 9th International Conference on Computer Supported Education*, 2017, pp. 173–183, doi: 10.5220/0006376501730183.

[3] A. Aninditya, M. A. Hasibuan, and E. Sutoyo, "Text mining approach using TF-IDF and naive bayes for classification of exam questions based on cognitive level of bloom's taxonomy," in *2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*, Nov. 2019, pp. 112–117, doi: 10.1109/IoTaIS47347.2019.8980428.

[4] B. Thomas and J. Chandra, "Random forest application on cognitive level classification of E-learning content," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 4, pp. 4372–4380, Aug. 2020, doi: 10.11591/ijece.v10i4.pp4372-4380.

[5] F. Nafa, S. Othman, and J. Khan, "Automatic concepts classification based on bloom's taxonomy using text analysis and the naïve Bayes classifier method," in *Proceedings of the 8th International Conference on Computer Supported Education*, 2016, pp. 391–396, doi: 10.5220/0005813303910396.

[6] K. Anekboon, “Feature selection for bloom’s question classification in thai language,” in *Advances in Intelligent Systems and Computing*, Springer International Publishing, 2019, pp. 152–162, doi: 10.1007/978-3-030-01174-1_12.

[7] R. Ahmad and W. G. Lutters, “Promoting reflective learning: the role of blogs in the classroom,” in *Online Communities and Social Computing*, Springer Berlin Heidelberg, 2011, pp. 3–11, doi: 10.1007/978-3-642-21796-8_1.

[8] E. A. Rahman Sidek and M. Md.Yunus, “Students’ experiences on using blog as learning journals,” *Procedia-Social and Behavioral Sciences*, vol. 67, pp. 135–143, Dec. 2012, doi: 10.1016/j.sbspro.2012.11.314.

[9] B. Montero-Fleta and C. Pérez-Sabater, “A research on blogging as a platform to enhance language skills,” *Procedia-Social and Behavioral Sciences*, vol. 2, no. 2, pp. 773–777, 2010, doi: 10.1016/j.sbspro.2010.03.100.

[10] T. Koraza, “What are the most popular topics on Medium?,” *Madx*, 2021. https://www.madx.digital/learn/the-most-popular-topics-on-medium (accessed Mar. 21, 2022).

[11] “Every idea needs a Medium.” *Medium*. https://medium.com/about (accessed May 26, 2021).

[12] B. S. Bloom, Ed., *Taxonomy of educational objectives: The classification of educational goals*. New York, Longmans, Green, 1956.

[13] D. R. Krathwohl, “A revision of bloom’s taxonomy: an overview,” *Theory Into Practice*, vol. 41, no. 4, pp. 212–218, Nov. 2002, doi: 10.1207/s15430421tip4104_2.

[14] A. Demetriou, S. Kazi, N. Makris, and G. Spanoudis, “Cognitive ability, cognitive self-awareness, and school performance: From childhood to adolescence,” *Intelligence*, vol. 79, Mar. 2020, doi: 10.1016/j.intell.2020.101432.

[15] K. Shah, H. Patel, D. Sanghvi, and M. Shah, “A comparative analysis of logistic regression, random forest and KNN models for the text classification,” *Augmented Human Research*, vol. 5, no. 1, Dec. 2020, doi: 10.1007/s41133-020-00032-0.

[16] C. C. Aggarwal and C. Zhai, “A survey of text classification algorithms,” in *Mining Text Data*, Boston, MA: Springer US, 2012, pp. 163–222, doi: 10.1007/978-1-4614-3223-4_6.

[17] H. Kim, J. Kim, J. Kim, and P. Lim, “Towards perfect text classification with Wikipedia-based semantic Naïve Bayes learning,” *Neurocomputing*, vol. 315, pp. 128–134, Nov. 2018, doi: 10.1016/j.neucom.2018.07.002.

[18] R. Arreerard and T. Senivongse, “Thai defamatory text classification on social media,” in *2018 IEEE International Conference on Big Data, Cloud Computing, Data Science and Engineering (BCD)*, Jul. 2018, pp. 73–78., doi: 10.1109/BCD2018.2018.00019.

[19] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, “Deep learning-based text classification,” *ACM Computing Surveys*, vol. 54, no. 3, pp. 1–40, Apr. 2022, doi: 10.1145/3439726.

[20] C. Piyaphakdeesakun, N. Facundes, and J. Polvichai, “Thai comments sentiment analysis on social networks with deep learning approach,” in *2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Jun. 2019, pp. 1–4, doi: 10.1109/ITC-CSCC.2019.8793324.

[21] K. Pasupa and T. S. N. Ayutthaya, “Hybrid deep learning models for Thai sentiment analysis,” *Cognitive Computation*, vol. 14, no. 1, pp. 167–193, Jan. 2022, doi: 10.1007/s12559-020-09770-0.

[22] M. E. Mavroforakis and S. Theodoridis, “A geometric approach to support vector machine (SVM) classification,” *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 671–682, May 2006, doi: 10.1109/TNN.2006.873281.

[23] F. Pedregosa *et al.*, “Scikit-learn: machine learning in python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[24] C. Haruechaiyasak, S. Kongyoung, and M. Dailey, “A comparative study on Thai word segmentation approaches,” in *2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, May 2008, pp. 125–128, doi: 10.1109/ECTICON.2008.4600388.

[25] W. Phatthiyaphaibun, “newnewthaicut,” *Github*, 2020. https://github.com/wannaphong/newnewthaicut (accessed May 28, 2021).

[26] “Chapter 2 let’s create a database for our application (in Thai),” *Medium*. https://medium.com/blogs-194/มาสร้าง-database-ให้-application-ของเรากันเถอะ-b5bf778bea (accessed May 26, 2021).

BIOGRAPHIES OF AUTHORS

Chalothon Chootong is a lecturer at the Kasetsart University, Sriracha campus. She received the Ph.D. in Computer Science from National Central University, Taiwan, the M.Sc. degrees in Computer Science from National Institute of Development Administration (NIDA), Thailand. Her research interests include social learning, data mining, recommender system, machine learning and AI. She can be contacted at email: chootong.c@ku.th.

Jirawan Charoensuk is a lecturer at the Kasetsart University, Sriracha campus. She received the Ph.D. in Computer Science from National Institute of Development Administration (NIDA), Thailand, the M.Eng. degrees in Computer Engineering from Kasetsart University, Thailand. Her research interests include text mining, natural language processing, machine learning and AI. She can be contacted at email: jirawan.charo@ku.th.