# Information Systems for Cultural Tourism Management Using Text Analytics and Data Mining Techniques

https://doi.org/10.3991/ijim.v16i09.30439

Thanet Yuensuk[1], Potsirin Limpinan[1],
Wongpanya Sararat Nuankaew[1], Pratya Nuankaew[2](✉)
[1]Faculty of Information Technology, Rajabhat Maha Sarakham University,
Maha Sarakham, Thailand
[2]School of Information and Communication Technology, University of Phayao,
Phayao, Thailand
pratya.nu@up.ac.th

**Abstract—**Using technology to deliver specific human interests is gaining attention. It results in humans being presented differently with what each individual wants. Therefore, this research aims to develop a culturally tourism recommended application using machine learning technology. It has three objectives: to develop a predictive model for cultural tourism management using text mining techniques, to evaluate the effectiveness of the cultural tourist attraction management model, and to assess the satisfaction of using the application for cultural tourism management. The research data was collected on Facebook conversations from 385 tourists (3,257 transactions) who had traveled to a famous tourist destination in Maha Sarakham Province. The prediction model development tools are three classification technique including Naïve Bayes, Neural Network, and K-Nearest Neighbor. The model performance evaluation tool consists of a confusion matrix and cross-validation methods. In addition, a questionnaire was used to assess the satisfaction of the application. The results showed that the model with the highest accuracy was modeled by Naïve Bayes technique with an accuracy of 91.65%. Simultaneously, the level of satisfaction with the application was high, with an average of satisfaction equal to 3.98 (S.D. equal to 0.69). It was therefore concluded that the application was accepted by it to be further expanded to offer more widespread research.

**Keywords—**cultural tourism management, opinion data mining, text mining, tourist attraction, tourist experience

## 1 Introduction

Chatbot technology is an artificial intelligence technology that is used to simulate providing information or answers to questions asked by users, regardless of whether they are in text or voice messages [1]–[3]. The working principle of this chatbot technology is powered by Artificial Intelligence (AI) by applying the principles of machine learning for analyzing user interactions [4]–[7]. In the selection of the most appropriate answer to the user's question. It combines natural language processing technology to translate computer language into a language that users can easily understand [8]. Moreover, the continuous improvements in big data and analysis using machine learning tools and the improved decision-making capabilities have resulted in the adoption of chatbot technology. Therefore, chatbot applications are becoming more and more popular [9], [10]. It is used to manage daily routines, provide necessary information through automated telephone systems, provide business information to inform products or services, as well as recommend the initial purchase of goods or services. However, the most popular chatbot technology is rule-based bots. The workflow is to create a Q&A rule. If a user asks a question, the bot will provide information that corresponds to the question, and there will be a troubleshooting process according to the rules set by the developers.

In the ecotourism dimension, the researchers found that the problem of travelers' desire for attractions was inconsistent with their experiences and backgrounds. As a result, traveling to various places is unhappy and unsettling. For this reason, the primary goal of the chatbot technology implementation in this research is to provide automated communication services where users can access and interact with chatbots through existing platforms. Please note that today's chatbot technology is not a replacement for all human conversations, especially if the conversation is very complicated [1], [2], [10]. This point is also considered a limitation of chatbot technology. However, if technology is continuously developed, it will result in the application of technology more efficiently. Therefore, this research has an important goal of developing a cultural tourism recommended application using machine learning technology. There are three objectives. The first objective is to develop a predictive model and the recommended application for cultural tourism management using text mining techniques. The second objective is to evaluate the effectiveness of the cultural tourist attraction recommendation model. Finally, the last objective is to assess the satisfaction of using the recommended application for cultural tourism management.

The research methodology based on the development of data mining; it is known as CRISP-DM: CRoss-Industry Standard Process for Data Mining [11]–[13]. There are six steps: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The research data is compiled from 385 tourists who have traveled in a famous tourist attraction in Maha Sarakham Province. The computation and collection of research samples were enumerated in the population and sample selection section. The data collected from tourists is a questionnaire on attitudes towards tourism in a particular location. In addition, the research framework and methodological concepts are presented in Figure 1.

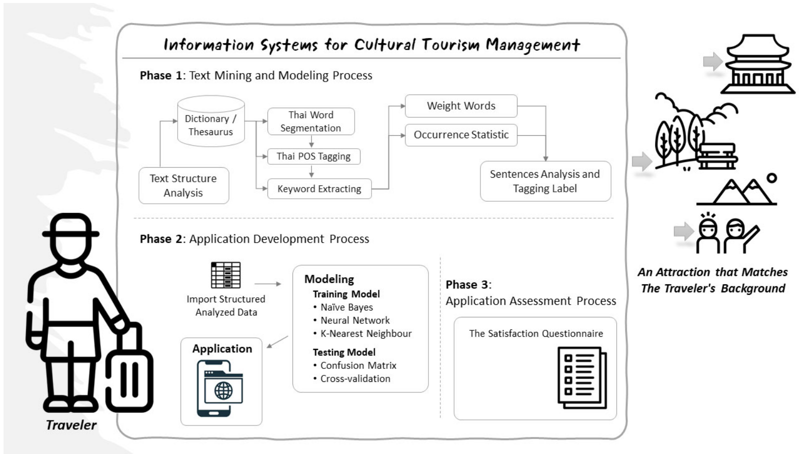


**Fig. 1.** The research framework and methodological concepts

Figure 1 presents the research framework and methodological concepts. It consists of three key phases. The first phase is the process of text analysis, which is an analysis of the traveler's experiences and backgrounds. The second phase is the process of model and application development. Lastly, the third phase is the process of model evaluation and application assessment, which aims to assess model performance and application satisfaction. Nevertheless, the research content presentation structure is divided into six sections: the first section is the presentation of the introduction and the research background. The second section presents the links of other research and related theories. The third section describes the research process that is divided into six steps based on the principles of data mining development. The fourth section is a summary of the research findings that have emerged. The fifth section is a discussion of the research findings and the outcomes obtained from the research. At the end, the last section is the conclusion. It provides a summary of the gist and recommendations derived from this research. Finally, the researchers have great confidence and hope that this research will be of the public interest with the hope that the research will be accepted, and the research results will be extended in the future.

## 2 Literature reviews

### 2.1 Tourists' attitudes towards technology

Understanding tourists' attitudes towards technology is an important area of tourism research as it relates to the interaction between the ability to use technology and the tourist experience, a dream destination that influences sustainable tourism development [14]. Technology is therefore being used as a tool to present alternatives to tourists [1], [2], [15]–[17]. An interesting examples are the use of technology to offer themes and attractions that match an individual's personality [15], using big social data for analysis of tourist behavior [18], a study of factors and influences on tourists [14], [19], and so on.

In addition, the use of technology in tourism is trending around the world [1], [15], [20], [21].

It underscores that technology is part of the travel apparatus, where the dimension of technology adoption consists of three dimensions being awakened. The first dimension is building a smart tourism city by integrating technology networks. It consists in the use of technology in tourist activities such as online vending, smart security services networking, improved transport services, linguistic services, smart city robots to guide visitors [5], [6], [22], [23]. The second dimension corresponds to the first dimension. It is the use of massive data to conduct analysis on the question of how big data technologies can improve tourism [4], [24]. Whilst smart city tourism is affected and benefits from the various sporting and entertainment activities that take place in the city at different times. Thousands of people are accepting and interested in participating in these events, although the types of events are diverse. For this reason, smart city tourism technology needs to be consistent and transferable, with event managers having to safely manage crowds and maximize event revenue. The last dimension is creating awareness to create an experience for tourists. In this regard, technology can be involved using augmented reality (AR) technology [25]–[27]. It is used to provide information or to offer directions. For example, when tourists use the Maps application they can search for landmarks and when they point their camera at that landmark while viewing the screen, they can choose to view information or historical images overlaid with the current scene.

From the literature review, it was found that understanding tourists' attitudes is an important question in the analysis for designing technology that is suitable for tourists. It is therefore sensible to study the context and attitudes of tourists in which the study and research process in this research is carried out in a scientific process.

### 2.2 Big data and innovation in tourism

Big data is an innovation of technology development based on information and communication technology. It is generated by referencing multiple datasets from different sources. It is supported by advanced data storage, analysis, and processing technologies. Beyond that, big data thinking is an opportunity based on the exponential growth of data volumes and the generation of unstructured data. With an unlimited amount of data, it is possible to study and discover patterns in the data. The number of innovations in using data to support the environment for tourism has emerged [18], [22], [28]. It features text-mining analysis for predictions in a variety of ways, including topic extraction, text classification, sentiment analysis, text clustering, and so on [28]. It is also used in the model of decision-making and application development such as tourist recommendation system, tourist satisfaction model, and AI technology-based service [9], [22], [27], [28].

For Thailand, part of the major income comes from tourism, with Thailand defining tourism as an important asset of the country. It appears research is bringing innovation and big data to drive research [29]–[31]. Consequently, these have inspired researchers to study and develop this research. Furthermore, the relationship perspective from another research that affects the researchers is that researchers want to apply accepted machine learning technology to study the behavior of tourists in Thailand.

Hereby, the issues and procedures for conducting research were presented in the research methodology section.

## 3 Research methodology

The research methodology consists of five key elements: Determination of the population and the sample, Development of research instruments, Data Collection, Model Generation and Efficiency, Using Statistics to Analyze Application Satisfaction. The researcher has taken the following actions.

### 3.1 Determination of the population and the sample

The population in this research was tourists in Maha Sarakham Province who had experience of traveling in cultural attractions in Maha Sarakham Province between 2021–2022. Please note that tourism data has not been collected between 2021–2022. Therefore, the researcher assigned an unknown population to calculate the sample [32], [33].

A common goal of social science research is to collect data representative of a population. A subset of the population is selected to represent the entire population in a study as a sample. Generally, there are several methods of classification in determining the sample size of the study for known and unknown population size. This study employed Cochran's technique in determining the sample size. The formula used to calculate the sample size for continuous data and unknown population size is n = ((z)2 * (σ) 2)/($e$)2 where, n is the sample size, $z = z$ value at reliability level or significance level (Reliability level 95% or significance level 0.05; $z = 1.96$, Reliability level 99% or significance level 0.01; $z = 2.58$), σ = standard deviation of the population, $e$ = acceptable sampling error ≈ (±5%) [If σ is unknown, defined $e$ as % of σ such as 8% of σ ($e = 0.08\sigma$) or 10% of σ ($e = 0.10\sigma$)].

Given that the largest acceptable discrepancy (e) is 10% of the population standard deviation and a statistical significance level of 0.05 for an unknown population, it was able to calculate the equivalent of 385 samples used in the research.

### 3.2 Development of research instruments

The research instruments consisted of three tools. The first tool is a chatbot model for cultural tourism management in Maha Sarakham Province. The second tool is a chatbot application for cultural tourism management in Maha Sarakham Province. The first two parts are detailed in the model generation and efficiency section. The last tool is the satisfaction questionnaire. It was used to assess the satisfaction of using the chatbot application by a sample of 30 students from Rajabhat Maha Sarakham University. Selected by a simple random sampling method as students in the Faculty of Information Technology, Rajabhat Maha Sarakham University. It consists of eight assessment issues as shown in Table 5.

### 3.3 Data collection

Data collection is classified into two parts. The first part is to collect data for modeling purposes. The data collection at this stage involves bringing information about the tourists' conversations through Facebook Messenger. It contains data from 385 Facebook accounts with 3,257 conversations. This data occurred in Phase 1: Text Processing to Analyze Traveler Experiences and Backgrounds and was used in Phase 2: Modeling and Application Development as shown in Figure 1.

The second data collection takes place for application assessment purposes. It is to collect information from 30 students in the Faculty of Information Technology, Rajabhat Maha Sarakham University.

### 3.4 Model generation and efficiency

Model development and efficiency were carried out according to a six-step data mining principle known as CRISP-DM: CRoss-Industry Standard Process for Data Mining. Additionally, there are two major phases in this step involved in the development of data mining principles: the first phase is text processing to analyze the traveler's experience and background as shown in Figure 1. It consists of three stages of CRISP-DM: Business Understanding, Data Understanding, and Data Preparation. The second phase is modeling and application development. It consists of three stages of CRISP-DM: Modeling, Evaluation, and Deployment. The details of each step are as follows.

**Business understanding.** In traditional human inquiries regarding the tourist cultural attractions in Maha Sarakham Province, there may be delayed or inconsistent answers to questions. It affects tourists from boredom and misinformation. Moreover, the vast amount of tourism information is scattered among the various departments of the size of the collection and coordination. These are the problems that have made researchers realize the importance that modern technology can solve.

Recognizing the importance of research, the researchers developed an application to guide the management of cultural tourism using text mining analytics and using machine learning technology to develop the application.

**Data understanding.** Understanding data can be said to be understanding the origin of the data or its form. Data today consists mainly of unstructured data formats, it manifests itself in the form of text and images [28], [34]. The major problem with most tourist inquiries is not the form of a sentence. It creates misunderstandings and confuses the reader and thus feeds back the wrong information. Some of the unstructured data the researchers found appeared in online conversations. The researchers then analyzed these data using text mining analysis techniques collected from 385 Facebook user accounts.

The Facebook Messenger question set data collection consists of eight question sets: (1) beverage shops and restaurants, (2) costs and fees for visiting, (3) facilities and data centers, (4) identity of community, (5) local wisdom, (6) location, (7) public transport, and (8) route. This data manipulation is in the data mining analysis phase, which is in the data preparation section.

**Data preparation.** Data preparation is the first phase as shown in Figure 1. It consists of five processes. The first step is to collect the questions and analyze the messages. This step explores how to classify questions using text mining techniques. It uses text mining processes and techniques to solve text classification problems. The goal is to cut and classify to identify key features for modeling. The data were then analyzed and modeled on the classification technique that gave the best classification performance. The second step is the word cutting stage. This research uses Python and Google Collab, which is an Open-Source program that has a process of cutting Thai words. It is then used to index keywords to exclude unimportant parts. The third step is the elimination of the stop word. Word types that are eliminated include prepositions, conjunctions, pronouns, adverbs, and interjections. The fourth step is to select and manage the question category feature. The fifth step is to create a keyword index. This research indexes the keywords in TF-Weighting format by substituting the text in the form of word vectors which calculate the weights for the index.

After completing the five-step process, it obtained the data ready for model development. The methods and tools that will be used to develop the model are discussed in the following sections.

**Modeling.** The text analysis modeling of this research was based on three machine learning techniques. It consists of Naïve Bayes technique, Neural Network technique, and K-Nearest Neighbors (K-NN) technique [35]–[37]. These three techniques are popular in the classification of supervised learning. The working principles and benefits of these three techniques are as follows:

Naïve bayes classifier is a famous model for classification based on the Bayes Rule. The advantage of the Naïve bayes classifier is that it is a model that is easy to understand and can also be easily evaluated. It can work quickly. If the conditional independence hypothesis persists and it can produce excellent results.

A neural network is a set of algorithms that attempt to recognize and relate hidden relationships in a dataset through a process that mimics the way the human brain works. Neural networks can adapt to changing information. Therefore, the network produces the best results without redesigning the export criteria. Its main assembly consists of three parts: an input layer, a processing layer, and an output layer. The input layer may be weighted according to various criteria. While the processing layer is hidden from view, it is responsible for the connections between these nodes that are like neurons. Finally, the output layer serves to display or show signs of various human activities.

The K-Nearest Neighbors (K-NN) algorithm is one of the techniques of machine learning algorithms that are easy to use, and the results are easy to understand. It can be used to effectively solve classification and regression problems. The K-NN algorithm uses the principle of near-similar comparison. In other words, similar things are close to each other. The working principle of K-NN therefore uses a method for calculating the distance of the object or class itself.

After the data has been prepared and the tools have been prepared, the next step is to test the developed model in which the methods are presented in the next section.

**Evaluation.** The model performance testing tool used for this research consisted of two tools. The first tool is a method of dividing the data to test the model known as “Cross-validation method”. Its principle is to divide data into two parts. The first part is used to create the model which is called “Training Dataset”. The rest of the data is used to test the model called “Testing Dataset”. In addition, the divided data is defined as intervals called “k-Fold”.

The second tool to analyze model performance is known as "Confusion Matrix Performance". There are three criteria for determining the model's performance. The first criterion is accuracy. It is used to represent the overall accuracy of the model, expressed as a percentage. The second criterion is the precision. It is used to show the precision of each answer (class), expressed as a percentage. The third criterion is recall. It is used to show the accuracy of each answer (class), expressed as a percentage. All criteria calculations are shown in Figure 2.

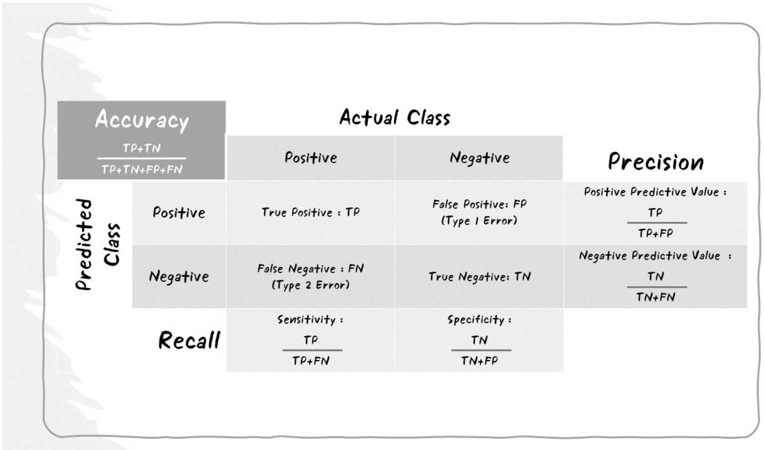


**Fig. 2.** The confusion matrix performance calculations

**Deployment.** The process of deploying a chatbot to recommend cultural attractions in Maha Sarakham Province is divided into two major steps. The first step is to prepare the chatbot server. This research used Chat Fuel's server and platform features to build a chatbot using natural language processing (NLP) based on machine learning techniques. It has a feature that helps in understanding the intent and entity of users in conversation. The second step is to connect the Facebook Messenger APIs. This process requires a tool by Facebook called "Facebook for Developers". This tool is what developers use to create applications that interact with Facebook. The communication process and the chatbot working process is shown in Figure 3.

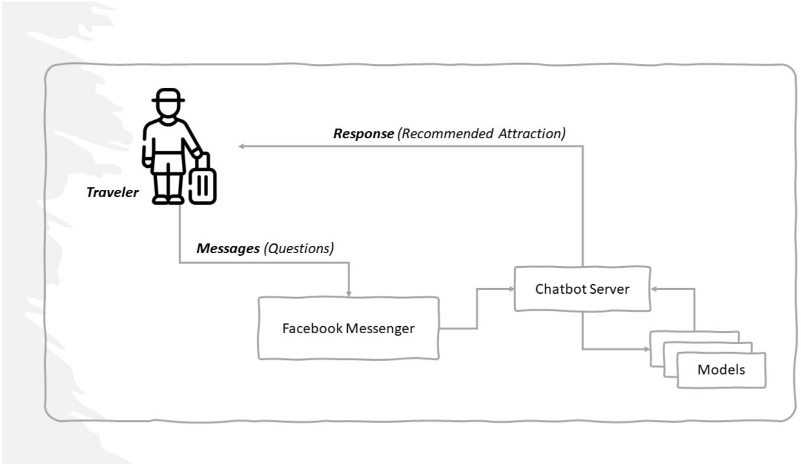


**Fig. 3.** The communication process and the chatbot working process

Figure 3 shows the process of using the chatbot application. When tourists enter questions into Facebook's chatbot application, the questions are then analyzed through a model and feedback is provided back to the tourists. After the development of the chatbot application was completed, the researchers continued to test it with a target audience of 30 samples.

### 3.5 Using statistics to analyze application satisfaction

The statistics used in this data analysis were the basic statistics to analyze user satisfaction with the chatbot application to recommend cultural attractions in Maha Sarakham Province. The tools used are frequency calculation, mean calculation, standard deviation calculation, and percentage calculation.

The satisfaction assessment criteria consisted of 5 levels. Level 1 has a rating of 1, which means extremely dissatisfied. Level 2 has a rating of 2, which means dissatisfied. Level 3 has a rating of 3, which means moderately satisfied. Level 4 has a rating of 4, which means satisfied. Level 5 has a rating of 5, which means extremely satisfied.

Interpretation is divided into 5 levels as follows: The mean is between 1.00 to 1.80. It can be interpreted as strongly disagree. The mean is between 1.81 to 2.60. It can be interpreted as disagree. The mean is between 2.61 to 3.40. It can be interpreted as neither agree nor disagree. The mean is between 3.41 to 4.20. It can be interpreted as agree. The mean is between 4.21 to 5.00. It can be interpreted as strongly agree.

## 4 Research results

The results of the study and research on the chatbots application for cultural tourism management were performed and the data were analyzed to determine the prototype model and application satisfaction. The issues are summarized as follows.

### 4.1 Predictive model for cultural tourism management

The project's research findings compiled a collection of general tourist questionnaires. It consists of frequently asked questions by tourists that are in the form of conversational sentences, not in the database format. It is then taken through the data preparation process, once the formatted data is obtained, the structured data is ready for modeling. In this process, the Thai Word Segmentation program is used to remove irrelevant words. The prepared data enters the model development process, which consists of three techniques: Naïve Bayes, Neural Network, and K-Nearest Neighbors. All the models tested were then taken through the Cross-Validation Method and Confusion Matrix process to determine the model's performance. The performance test results for each model are as follows.

**Modeling with Naïve Bayes technique.** The modeling results of the Naïve Bayes technique applied by the classification results with precision, recall, and F1-score values are shown in Table 1.

**Table 1.** Performance analysis results with Naïve Bayes technique

| Class | Naïve Bayes Technique Performances | | |
|---|---|---|---|
| | **Precision** | **Recall** | **F1-Score** |
| (1) Beverage shops and restaurants | **94.90%*** | 85.05% | **93.47%*** |
| (2) Costs and fees for visiting | 89.36% | 87.62% | 90.81% |
| (3) Facilities and data centers | 83.60% | 92.08% | 84.72% |
| (4) Identity of community | 92.31% | **96.06%*** | 93.39% |
| (5) Local wisdom | 91.92% | 94.49% | 88.35% |
| (6) Location | 89.69% | 85.87% | 90.16% |
| (7) Public transport | 89.32% | 92.00% | 88.46% |
| (8) Route | 91.01% | 92.05% | 91.53% |

From Table 1 shows the results of the Naïve Bayes classification of question sets, it was found that the precision performance was the most 94.90% in the beverage shops and restaurants class. The recall performance was as high as 96.06% in the identity of community class. Lastly, the F1-score performance was as high as 93.47% in the beverage shops and restaurants class. The results of these performance tests were further compared with other models.

**Modeling with Neural Network technique.** The modeling results of the Neural Network technique applied by the classification results with precision, recall, and F1-score values are shown in Table 2.

**Table 2.** Performance analysis results with Neural Network technique

| Class | Neural Network Technique Performances | | |
|---|---|---|---|
| | Precision | Recall | F1-Score |
| (1) Beverage shops and restaurants | **96.91%*** | **94.57%*** | 91.26% |
| (2) Costs and fees for visiting | 92.22% | 88.12% | 88.77% |
| (3) Facilities and data centers | 87.82% | 86.24% | 90.73% |
| (4) Identity of community | 89.55% | 94.26% | 89.55% |
| (5) Local wisdom | 92.55% | 89.55% | **93.55%*** |
| (6) Location | 83.33% | 93.84% | 88.46% |
| (7) Public transport | 91.75% | 85.57% | 89.90% |
| (8) Route | 95.18% | 89.77% | 92.40% |

From Table 2 shows the results of the Neural Network classification of question sets, it was found that the precision performance was the most 96.91% in the beverage shops and restaurants class. The recall performance was as high as 94.57% also in the beverage shops and restaurants class. Lastly, the F1-score performance was as high as 93.55% in the local wisdom class. The results of these performance tests were further compared with other models.

**Modeling with K-Nearest Neighbors technique.** The modeling results of the K-Nearest Neighbors technique applied by the classification results with precision, recall, and F1-score values are shown in Table 3.

**Table 3.** Performance analysis results with K-Nearest Neighbors technique

| Class | K-Nearest Neighbors Technique Performances | | |
|---|---|---|---|
| | Precision | Recall | F1-Score |
| (1) Beverage shops and restaurants | 92.55% | 89.55% | **92.55%*** |
| (2) Costs and fees for visiting | 88.55% | 89.26% | 87.55% |
| (3) Facilities and data centers | 93.91% | 91.57% | 91.26% |
| (4) Identity of community | 87.82% | 86.24% | 90.73% |
| (5) Local wisdom | 92.22% | 88.12% | 88.77% |
| (6) Location | 91.75% | 85.57% | 89.90% |
| (7) Public transport | **95.18%*** | 89.77% | 91.40% |
| (8) Route | 89.55% | **94.26%*** | 89.55% |

From Table 3 shows the results of the K-Nearest Neighbors classification of question sets, it was found that the precision performance was the most 95.18% in the public transport class. The recall performance was as high as 94.26% also in the route class. Lastly, the F1-score performance was as high as 92.55% in the beverage shops and restaurants class. The results of these performance tests were further compared with other models.

The results of the efficiency analysis in Table 1 to Table 3 show that each class has different performance from the results of each model. Where, Table 4 reports the results of the performance classification of various models by analyzing the accuracy performance and time of model development.

**Table 4.** Classification of model performance

| Classifiers | Accuracy | Times(s) |
|---|---|---|
| Naïve Bayes | 94.77% | 10 |
| Neural Network | 86.67% | 24 |
| K-Nearest Neighbors | 91.48% | 18 |

From Table 4 the results of the classification of question sets by three techniques were Naïve Bayes, Neural Network, and K-Nearest Neighbors classifiers. The results were compared with the three highest-accuracy techniques in which Naïve Bayes had the highest accuracy of 91.65% and it took the least time to develop the model at 54 seconds. Once an efficient and reasonable model is obtained. The next part is the implementation of the Facebook Messenger chatbot, where an example of its implementation is shown in Figure 4.

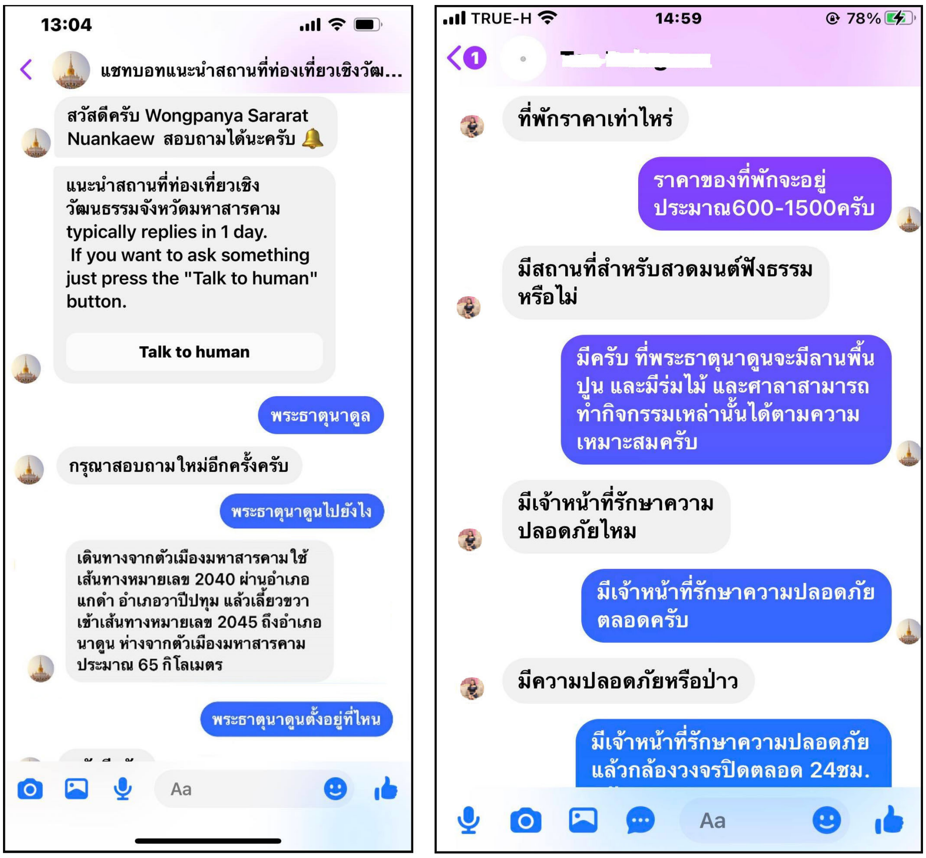


**Fig. 4.** Deploying the model to facebook messenger chatbot

Figure 4 shows an example of the model deployment in the Facebook Messenger chatbot program. The outcome of this application's system is based on a model developed with machine learning and all the performance tested to develop a chatbot system on Facebook Messenger that can answer questions automatically. After obtaining the application, it is tested with a target audience that has been defined.

### 4.2 Chatbot satisfaction assessment results

The results of satisfaction with the chatbot application for cultural tourism management in Maha Sarakham Province for each stage, which were presented as mean, standard deviation, and their interpretation are presented in Table 5. The sample group was 30 students of the Faculty of Information Technology, Rajabhat Maha Sarakham University.

**Table 5.** Chatbot satisfaction assessment results

| Stage | Satisfaction Level | | |
|---|---|---|---|
| | **Mean** | **S.D.** | **Interpretation** |
| 1. Satisfaction toward providing responses to questions via chatbots | 4.30 | 0.67 | Strongly agree |
| 2. Satisfaction toward providing appropriate and up-to-date information | 4.13 | 0.68 | Agree |
| 3. Satisfaction toward using appropriate language via chatbots | 4.02 | 0.69 | Agree |
| 4. Satisfaction toward the correctness of answering questions | 4.25 | 0.68 | Strongly agree |
| 5. Satisfaction toward the speed in responding to questions | 4.24 | 0.65 | Strongly agree |
| 6. Satisfaction toward a diverse set of responses | 4.23 | 0.62 | Strongly agree |
| 7. Satisfaction toward the friendliness of the system | **4.35** | 0.63 | Strongly agree |
| 8. Satisfaction toward the functionality and complexity of the system | 4.32 | 0.60 | Strongly agree |
| Overall Satisfaction: | 4.17 | 0.66 | Agree |

Table 5 presents the results of the user satisfaction assessment of 30 students. It was found that overall users had a high level of satisfaction with accepting the chatbot application for cultural tourism management. It has an average satisfaction rating of 4.17. Where the issue that was recognized as having the highest satisfaction was the satisfaction with the friendliness of the system. It has an average satisfaction rating of 4.35. However, other issues were accepted as well. It can be concluded that the developed application is accepted according to the target audience that has been defined.

## 5 Research discussions

The results of a research study on the development of an application for cultural tourism management in Maha Sarakham Province, in which the model development steps were determined according to the six-step CRISP-DM data mining process. As well as assessing the satisfaction of the application, it can be summarized as two main points.

### 5.1 Success in the development of predictive models

The first success is that the prediction model for cultural tourism management of Maha Sarakham Province can actually be developed within the specified scope. In the model development, three techniques were compared to select the most efficient model. All 3 techniques consist of Naïve Bayes, Neural Network, and K-Nearest Neighbors. The comparison results of the three models are as follows: The classification results of the question sets using Naïve Bayes technique, it was found that the class with the highest precision rating was the beverage shops and restaurants class, with a precision of 94.90%. In addition, the Recall performance was as high as 96.06% in the identity of community class. Lastly, the F1-Score performance was as high as 93.47% in the beverage shops and restaurants class as detailed in the Table 1. The classification results of the question sets using Neural Network technique, it was found that the class with the highest precision rating was the beverage shops and restaurants class, with a precision of 96.91%. In addition, the Recall performance was as high as 94.57% in the beverage shops and restaurants class. Lastly, the F1-score performance was as high as 93.55% in the local wisdom class as detailed in the Table 2.

Lastly, the classification results of the question sets using K-Nearest Neighbors technique, it was found that the class with the highest precision rating was the public transport class, with a precision of 95.18%. In addition, the Recall performance was as high as 94.26% in the route class. Lastly, the F1-score performance was as high as 92.55% in the beverage shops and restaurants class as detailed in the Table 3.

Moreover, the results of the analysis by classification according to the validity of the developed models are presented in Table 4. It was found that the models created using the Naïve Bayes technique had the highest accuracy and the least time to develop the model, which has been selected for use in Facebook's Messenger Chatbot. An example of the model deployment in the Facebook's Messenger Chatbot application is shown in Figure 4.

### 5.2 Success in application development

The developed application was tested to assess the satisfaction of testers with the use of the Facebook's Messenger Chatbot application. The application testers were selected samples from 30 students of the Faculty of Information Technology, Rajabhat Maha Sarakham University. A detailed summary of the eight assessments is presented in Table 5.

The results of the assessment of the chatbot application for the management of cultural tourism through the Facebook Messenger platform, it was found that the application testers had a high level of satisfaction. The overall satisfaction of the application testers was found to be an average of 3.98. It can mean that the tester accepts the use of the developed application. The point that the testers accepted the most was the satisfaction toward the correctness of answering questions issue. It had an average of 4.37 (S.D. equal to 0.45). The second point that the testers most accepted was the satisfaction toward providing appropriate and up-to-date information issue. It had an average of 4.10 (S.D. equal to 0.36). In addition, other issues have been accepted as well. It can be concluded that the applications developed are accepted in all dimensions.

## 6 Conclusion

Innovation is not just about creating something new; it should make current technology uphold the foundation of a national institution, which means a nation's community. Therefore, this research aims to develop innovations for communities and developing countries. The goal of this research is to bring university knowledge to the local community. Its main objective is to develop technology to enable communities to benefit from modern innovations. The sub-objectives of this research consist of three issues: to develop a predictive model for cultural tourism management using text mining techniques, to evaluate the effectiveness of the cultural tourist attraction management model, and to assess the satisfaction of using the application for cultural tourism management. The research data was collected on Facebook conversations from 385 tourists (3,257 transactions) who had traveled to a famous tourist destination in Maha Sarakham Province. It was found that informants were very pleased with the research objectives. The prediction model development tools are three classification technique including Naïve Bayes, Neural Network, and K-Nearest Neighbor. The model performance evaluation tool consists of a confusion matrix and cross-validation methods. Model development and model testing tools make it possible to discover unique models. It found the correlation of data and models that enable the system to deliver results to users with high satisfaction. While the questionnaire was used to assess the satisfaction of the application indicating acceptance of the innovation that had been created. The results showed that the model with the highest accuracy was modeled by Naïve Bayes technique with an accuracy of 91.65%. Simultaneously, the level of satisfaction with the application was high, with an average of satisfaction equal to 3.98 (S.D. equal to 0.69). The conclusions of the testers using the developed applications show that the researchers developed a responsive information system. It was therefore concluded that the application was accepted by it to be further expanded to offer more widespread research. The researcher has strong confidence that what has been studied in this research has been successful and will be accepted for future use.

## 7 Acknowledgment

This research project was supported by the Thailand Science Research and Innovation Fund and the University of Phayao (Grant No. FF65-UoE006). In addition, this research was supported by many advisors, academicians, researchers, students, staff, and agencies from two organizations: the School of Information and Communication Technology at the University of Phayao, and the Faculty of Information Technology at the Rajabhat Maha Sarakham University. The authors would like to thank all of them for their support and collaboration in making this research possible.

## 8 References

[1] G. Sperlí, "A cultural heritage framework using a deep learning based chatbot for supporting tourist journey," *Expert Syst. Appl.*, vol. 183, p. 115277, Nov. 2021. https://doi.org/10.1016/j.eswa.2021.115277

[2] M. Casillo, F. Clarizia, G. D'Aniello, M. De Santo, M. Lombardi, and D. Santaniello, "CHAT-Bot: a cultural heritage aware teller-bot for supporting touristic experiences," *Pattern Recognit. Lett.*, vol. 131, pp. 234–243, Mar. 2020. https://doi.org/10.1016/j.patrec.2020.01.003

[3] I. D. Wahyono *et al.*, "Shared nearest neighbour in text mining for classification material in online learning using mobile application," *Int. J. Interact. Mob. Technol. IJIM*, vol. 16, no. 04, Art. no. 04, Feb. 2022. https://doi.org/10.3991/ijim.v16i04.28991

[4] J. Romão, K. Kourtit, B. Neuts, and P. Nijkamp, "The smart city as a common place for tourists and residents: a structural analysis of the determinants of urban attractiveness," *Cities*, vol. 78, pp. 67–75, Aug. 2018. https://doi.org/10.1016/j.cities.2017.11.007

[5] G. Zhou, F. Kurauchi, S. Ito, and R. Du, "Identifying golden routes in tourist areas based on AMP collectors," *Asian Transp. Stud.*, vol. 8, p. 100052, Jan. 2022. https://doi.org/10.1016/j.eastsj.2021.100052

[6] M. Visan, S. L. Negrea, and F. Mone, "Towards intelligent public transport systems in Smart Cities; Collaborative decisions to be made," *Procedia Comput. Sci.*, vol. 199, pp. 1221–1228, Jan. 2022. https://doi.org/10.1016/j.procs.2022.01.155

[7] M. Paolanti *et al.*, "Tourism destination management using sentiment analysis and geo-location information: a deep learning approach," *Inf. Technol. Tour.*, vol. 23, no. 2, pp. 241–264, Jun. 2021. https://doi.org/10.1007/s40558-021-00196-4

[8] S. Lu, G. Li, and M. Xu, "The linguistic landscape in rural destinations: a case study of Hongcun Village in China," *Tour. Manag.*, vol. 77, p. 104005, Apr. 2020. https://doi.org/10.1016/j.tourman.2019.104005

[9] M. Li, D. Yin, H. Qiu, and B. Bai, "A systematic review of AI technology-based service encounters: implications for hospitality and tourism operations," *Int. J. Hosp. Manag.*, vol. 95, p. 102930, May 2021. https://doi.org/10.1016/j.ijhm.2021.102930

[10] S. Mohamad Suhaili, N. Salim, and M. N. Jambli, "Service chatbots: a systematic review," *Expert Syst. Appl.*, vol. 184, p. 115461, Dec. 2021. https://doi.org/10.1016/j.eswa.2021.115461

[11] M. Cazacu and E. Titan, "Adapting CRISP-DM for social sciences," *BRAIN Broad Res. Artif. Intell. Neurosci.*, vol. 11, no. 2Sup1, Art. no. 2Sup1, May 2021. https://doi.org/10.18662/brain/11.2Sup1/97

[12] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, Jan. 2021. https://doi.org/10.1016/j.procs.2021.01.199

[13] J. Venter, A. de Waal, and C. Willers, "Specializing CRISP-DM for evidence mining," in *Advances in Digital Forensics III*, New York, NY, 2007, pp. 303–315. https://doi.org/10.1007/978-0-387-73742-3_21

[14] K. Panwanitdumrong and C.-L. Chen, "Investigating factors influencing tourists' environmentally responsible behavior with extended theory of planned behavior for coastal tourism in Thailand," *Mar. Pollut. Bull.*, vol. 169, p. 112507, Aug. 2021. https://doi.org/10.1016/j.marpolbul.2021.112507

[15] C.-K. Lee, M. S. Ahmad, J. F. Petrick, Y.-N. Park, E. Park, and C.-W. Kang, "The roles of cultural worldview and authenticity in tourists' decision-making process in a heritage tourism destination using a model of goal-directed behavior," *J. Destin. Mark. Manag.*, vol. 18, p. 100500, Dec. 2020. https://doi.org/10.1016/j.jdmm.2020.100500

[16] E. M. Abuelrub and H. M. Solaiman, "A tourism E-guide system using mobile integration," *Int. J. Interact. Mob. Technol. IJIM*, vol. 4, no. 2, Art. no. 2, Mar. 2010. https://doi.org/10.3991/ijim.v4i2.1051
[17] C. Suphachaimongkol, C. Ratanatamskul, S. Silapacharanan, and P. Utiswannakul, "Development of mobile application for sustainable creative tourism assessment using confirmatory factor analysis approach," *Int. J. Interact. Mob. Technol. IJIM*, vol. 13, no. 06, Art. no. 06, Jun. 2019. https://doi.org/10.3991/ijim.v13i06.10500
[18] M. T. Cuomo, D. Tortora, P. Foroudi, A. Giordano, G. Festa, and G. Metallo, "Digital transformation and tourist experience co-design: big social data for planning cultural tourism," *Technol. Forecast. Soc. Change*, vol. 162, p. 120345, Jan. 2021. https://doi.org/10.1016/j.techfore.2020.120345
[19] S. Tse and V. W. S. Tung, "Understanding residents' attitudes towards tourists: connecting stereotypes, emotions and behaviours," *Tour. Manag.*, vol. 89, p. 104435, Apr. 2022. https://doi.org/10.1016/j.tourman.2021.104435
[20] S. (Sixue) Jia, "Motivation and satisfaction of Chinese and U.S. tourists in restaurants: a cross-cultural text mining of online reviews," *Tour. Manag.*, vol. 78, p. 104071, Jun. 2020. https://doi.org/10.1016/j.tourman.2019.104071
[21] J. Wen, S. (Sam) Huang, and T. Ying, "Relationships between Chinese cultural values and tourist motivations: a study of Chinese tourists visiting Israel," *J. Destin. Mark. Manag.*, vol. 14, p. 100367, Dec. 2019. https://doi.org/10.1016/j.jdmm.2019.100367
[22] A. Gutiérrez, A. Domènech, B. Zaragozí, and D. Miravet, "Profiling tourists' use of public transport through smart travel card data," *J. Transp. Geogr.*, vol. 88, p. 102820, Oct. 2020. https://doi.org/10.1016/j.jtrangeo.2020.102820
[23] D. Ilić, I. Milošević, and T. Ilić-Kosanović, "Application of unmanned aircraft systems for smart city transformation: case study Belgrade," *Technol. Forecast. Soc. Change*, vol. 176, p. 121487, Mar. 2022. https://doi.org/10.1016/j.techfore.2022.121487
[24] E. Sigalat-Signes, R. Calvo-Palomares, B. Roig-Merino, and I. García-Adán, "Transition towards a tourist innovation model: the smart tourism destination: reality or territorial marketing?," *J. Innov. Knowl.*, vol. 5, no. 2, pp. 96–104, Apr. 2020. https://doi.org/10.1016/j.jik.2019.06.002
[25] T.-L. Huang, "Restorative experiences and online tourists' willingness to pay a price premium in an augmented reality environment," *J. Retail. Consum. Serv.*, vol. 58, p. 102256, Jan. 2021. https://doi.org/10.1016/j.jretconser.2020.102256
[26] E. R. Fino, J. Martín-Gutiérrez, M. D. M. Fernández, and E. A. Davara, "Interactive tourist guide: connecting Web 2.0, augmented reality and QR codes," *Procedia Comput. Sci.*, vol. 25, pp. 338–344, Jan. 2013. https://doi.org/10.1016/j.procs.2013.11.040
[27] N. Chung, H. Han, and Y. Joun, "Tourists' intention to visit a destination: the role of augmented reality (AR) application for a heritage site," *Comput. Hum. Behav.*, vol. 50, pp. 588–599, Sep. 2015. https://doi.org/10.1016/j.chb.2015.02.068
[28] Q. Li, S. Li, S. Zhang, J. Hu, and J. Hu, "A review of text corpus-based tourism big data mining," *Appl. Sci.*, vol. 9, no. 16, Art. no. 16, Jan. 2019. https://doi.org/10.3390/app9163300
[29] H. A. T. Nguyen *et al.*, "Comparative carbon footprint assessment of agricultural and tourist locations in Thailand," *J. Clean. Prod.*, vol. 269, p. 122407, Oct. 2020. https://doi.org/10.1016/j.jclepro.2020.122407
[30] S. Fuktong *et al.*, "A survey of stereotypic behaviors in tourist camp elephants in Chiang Mai, Thailand," *Appl. Anim. Behav. Sci.*, vol. 243, p. 105456, Oct. 2021. https://doi.org/10.1016/j.applanim.2021.105456
[31] Y. Jeaheng and H. Han, "Thai street food in the fast growing global food tourism industry: Preference and behaviors of food tourists," *J. Hosp. Tour. Manag.*, vol. 45, pp. 641–655, Dec. 2020. https://doi.org/10.1016/j.jhtm.2020.11.001

[32] J. T. Roscoe, *Fundamental research statistics for the behavioral sciences*. New York: Holt, Rinehart and Winston, 1969. Accessed: Feb. 08, 2022. [Online]. Available: http://archive.org/details/fundamentalresea0000rosc
[33] J. Charan and T. Biswas, "How to calculate sample size for different study designs in medical research?," *Indian J. Psychol. Med.*, vol. 35, no. 2, pp. 121–126, Apr. 2013. https://doi.org/10.4103/0253-7176.116232
[34] L. Zhang, Z. Qi, and F. Meng, "A review on the construction of business intelligence system based on unstructured image data," *Procedia Comput. Sci.*, vol. 199, pp. 392–398, Jan. 2022. https://doi.org/10.1016/j.procs.2022.01.048
[35] R. Blanquero, E. Carrizosa, P. Ramírez-Cobo, and M. R. Sillero-Denamiel, "Variable selection for Naïve Bayes classification," *Comput. Oper. Res.*, vol. 135, p. 105456, Nov. 2021. https://doi.org/10.1016/j.cor.2021.105456
[36] A.-A. Tulbure, A.-A. Tulbure, and E.-H. Dulf, "A review on modern defect detection models using DCNNs – Deep convolutional neural networks," *J. Adv. Res.*, vol. 35, pp. 33–48, Jan. 2022. https://doi.org/10.1016/j.jare.2021.03.015
[37] J. R. Rico-Juan, J. J. Valero-Mas, and J. Calvo-Zaragoza, "Extensions to rank-based prototype selection in k-Nearest Neighbour classification," *Appl. Soft Comput.*, vol. 85, p. 105803, Dec. 2019. https://doi.org/10.1016/j.asoc.2019.105803

## 9 Authors

**Thanet Yuensuk** is currently an instructor at the Faculty of Information Technology, Rajabhat Maha Sarakham University, Maha Sarakham, 44000, Thailand. (Email: thanet.yu@rmu.ac.th) His research interests are learning media development, knowledge management, information systems development, and information technology management.

**Potsirin Limpinan** is currently an assistant professor at the Faculty of Information Technology, Rajabhat Maha Sarakham University, Maha Sarakham, 44000, Thailand. (Email: potsirin.li@rmu.ac.th) Her research interests are learning media development, knowledge management, information systems development, and information technology management.

**Wongpanya Sararat Nuankaew** is currently an assistant professor at the Faculty of Information Technology, Rajabhat Maha Sarakham University, Maha Sarakham, 44000, Thailand. (Email: wongpanya.nu@rmu.ac.th) Her research interests are digital education, innovation and knowledge management, data science, and big data and information technology management.

**Pratya Nuankaew** is currently an instructor at the School of Information and Communication Technology, University of Phayao, Phayao, 56000, Thailand. (Email: pratya.nu@up.ac.th) He is the corresponding author on this research. His research interests are applied informatics technologies, behavioral sciences analysis with technologies, computer-supported collaborative learning, data science in education, educational data mining, learning analytics and learning styles, learning strategies for lifelong learning, self-regulated learning, social network analysis, and ubiquitous computing.

Article submitted 2022-02-26. Resubmitted 2022-03-27. Final acceptance 2022-03-29. Final version published as submitted by the authors.