Analyzing and Tracking Student Educational Program Interests on Social Media with Chatbots Platform and Text Analytics

https://doi.org/10.3991/ijim.v17i05.31593

Patchara Nasa-Ngium[1], Wongpanya Sararat Nuankaew[2], Pratya Nuankaew[3](✉)
[1] Faculty of Science and Technology, Rajabhat Maha Sarakham University, Maha Sarakham, Thailand
[2] Faculty of Information Technology, Rajabhat Maha Sarakham University, Maha Sarakham, Thailand
[3] School of Information and Communication Technology, University of Phayao, Phayao, Thailand
pratya.nu@up.ac.th

Abstract—This research presents a chatbot application to provide educational information for university students. There are three objectives: 1) to study the problem of providing information to university students with chatbots, 2) to develop a model and construct a chatbot to predict the interest of university students, and 3) to assess the satisfaction of the information provided by the chatbot application. The research datasets were the conversations from the Messenger Facebook Page of the Faculty of Information Technology, Rajabhat Maha Sarakham University, during the academic year 2020-2021. In total, there were 1,094 transactions used in this research work. Furthermore, data mining and machine learning techniques, including CRISP-DM, Naïve Bayes, K-Nearest Neighbors, and Neural Network, were used as the research tools. The cross-validation and confusion matrix techniques were used to test the model performance. Moreover, a questionnaire was the application satisfaction assessment tool for 30 respondents. As a result, it showed that the developed model provided high-level results, which are 88.73% accuracy and an average of 3.97 for application satisfaction. In the future, the researchers plan to apply the results for the next academic year and expand into other academic programs.

Keywords—applied informatics, text analytics, educational data mining, eruptive technology, technology-enhanced learning

1 Introduction

Organizational communication for providing information about educational programs to university students is a fundamental problem for motivating admission to study programs. Students need communication that is speedy, snappy, and up to date. Using personnel as a primary communication tool can be problematic, for example, delays, inaccuracies of information, and inaccurate information. In a technological era

where people prefer communicating via smartphones as the main channel, chatbots have steadily become a popular medium for interacting between agencies and users. It increases its popularity and acceptance endlessly [1]–[6]. Advancements in Artificial Intelligence (AI) and machine learning technologies have empowered chatbots to become powerful tools and technologies that continue playing an ever-increasing role and influencing human communications [3]. Moreover, chatbots have been applied for a wide range of different purposes to build customer relationships and promote educational programs of educational institutions [1], [4] through commercial, public relations [5], and corporate communications.

Due to the COVID-19 pandemic, universities will have to organize online classes to solve problems [7]–[15]. For the Bachelor of Information Technology at the Faculty of Information Technology, Rajabhat Maha Sarakham University, the adoption of chatbot technology in universities to solve the problem of misinformation for students has been ignored. It interrupts communication between students and educational institutions. Students often get delayed information, and many other problems affect learners. So, the research question is, how do educational institutions provide tools and strategies for communicating with learners to meet the need for appropriate information for the students in their educational institutions? This research question anticipates determining the behavioral patterns of learners interested in further education and their needs.

Consequently, the researchers intend to study and support education with three objectives. The first objective was to study the problem of providing information to university students with chatbots. This objective is to promote the learning and adoption of chatbot technology in the education industry. The second objective was to develop a model and construct a chatbot to predict the interest of university students. This objective focus on piloting and stimulating further learning of the implementation of chatbot technology in the education sector. The last objective is to assess the satisfaction of the information provided by the chatbot application. The application satisfaction assessment aims to build user awareness and identify flaws for future improvements.

The researchers made two crucial research hypotheses and beliefs.

- H1: Using artificial intelligence and text mining technologies can effectively develop a model for predicting interest and providing information for students effectively.
- H2: Model prototypes and applications that have been tested with scientific processes will garner a high level of user satisfaction.

An overview of the methodological research relationship is presented in Table 1. The conceptual research framework is divided into three phases. The first phase is the collection of data by applying text analytics technology. Initially, the data was collected as unstructured data, and it is necessary to manage it in a structured data format that prepares the data of the research methodology. The second phase is the model development phase. In this section, the researchers selected essential tools to analyze the outcomes for predicting admissions interest to the Bachelor of Information Technology at the Faculty of Information Technology, Rajabhat Maha Sarakham University. It consists of three predictive model development tools: Naïve Bayes, K-Nearest Neighbors,

and Neural Network. It also consists of two testing techniques, including a cross-validation method and a confusion matrix technique. The second phase of this research relies on modeling and evaluating the research methodology, while the third phase is an application development based on the research deployment methodology. The last phase consists of two stages, including the application development and the application satisfaction assessment stage.

Table 1. Research Methodological Relationship

Research Objectives	Research Processes	CRISP-DM Stages
(1) To study the problem of providing information to university students with chatbots	(1) Define research problems (2) Formulate research hypothesis	(1) Business Understanding
	(3) Collect data	(2) Data Understanding
	(4) Convert unstructured data into structured data (5) Label data set	(3) Data Preparing
(2) To develop a model and construct a chatbot to predict the interest of university students	(6) Develop prototype models	(4) Modeling
	(7) Evaluate the performance of prototype models (8) Select the most reasonable model	(5) Evaluation
(3) To assess the satisfaction of the information provided by the chatbot application	(9) Develop an application prototype (10) Find satisfaction with the application (11) Summary and improvement	(6) Deployment

Table 1 provides the relationship of the research development process. It consists of three research objectives, including eleven research processes and six functional areas based on the CRISP-DM Data Mining Development Principles. Besides, this research designed a six-section of presentation outline. The first section is an introduction, which presents the significance of the research. The second section is the literature reviews and related works, which summarize the existing research that influenced this research. The third section is the research methodology based on the principle of data mining development using the CRISP-DM technique [16]–[18], consisting of six stages.

The fourth section is the research findings, which classify the reports according to research objectives. The fifth section discusses the results, which analyzes the observations and findings from the research study. The last section is the conclusion section. It consists of comparing research findings against objectives, summarizing the pros and cons, and applying them to further improvements in future work.

2 Literature reviews and related works

This review aims to initiate understanding and raise awareness among readers. This section details the effectiveness of the educational chatbot application. A chatbot or chatbot application is a software application that can operate automatically through pre-

defined operating conditions. A chatbot application is a program created to simulate conversations with users in natural language through a multi-platform messaging application [19]. The outstanding feature of the chatbot application is that it does not require installation on the device, and it is therefore popular as a communication tool with learners [19]–[22].

Chatbots are increasingly being used for education to improve student interactions by relying heavily on online platforms for students' communication and many other activities [4], [23]. Prime examples of educational chatbot applications are teaching, administrative, assessment, advisory, and scholarly research activities [4]. Although the educational process is an integral part of human development, the adoption of technology such as chatbots appears in small numbers. It seems that research reports on improving the quality of education are limited. Moreover, educational technology developments have focused on building tools and studying the learning process rather than developing applications to help design problem-solving in the education system [1], [24], [25].

Nevertheless, researchers' perspective believes it lacks the adoption of modern technology, including artificial intelligence technology, to support the education system. Strong ideas of researchers are that the design of student success needs to be interpreted in the interest and acceptance of the learners themselves. This belief base has been accepted [26]–[29]; however, it is essential to prioritize the development of learners' achievement in the education system due to many weaknesses.

3 Research methodology

The core of this research methodology is based on CRISP-DM data mining development principles [16]–[18], [30]. The central part is related to six elements: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. All six aspects work concerning the research process where the researchers show the relationship between research objectives, research development processes according to the research framework, and CRISP-DM data mining development principles, as shown in Table 1.

3.1 Business understanding

In conducting research, it is imperative to understand the purpose of using the data to analyze the research problem, which is the concept of business understanding [16], [17]. In this section, the researchers set the work direction in two main areas: formulating research questions and hypotheses, as presented in the introduction.

3.2 Data understanding

Data Understanding is the process of creating awareness and communicating with research questions [16], [17]. It creates an understanding of data by collating relevant data, conducting a selection of important data, and managing the suitability of data in

terms of quantity and data adequacy. The element of data understanding consists of four aspects including initial data collection, data description, data exploration, and data quality verification. This research followed the four aspects mentioned.

In the initiation of data collection, the researchers determined the period for data collection from inquiries between interested parties in the Information Technology Program and the Facebook Page of the Faculty of Information Technology determined during the academic year 2020-2021. In the process of the data description, the researchers scoped the data synthesis by communicating issues that related to the student's interest in further education. It involves exploring the data used to analyze the relevant question of a research problem. In a final step, the researchers proceeded to confirm the quality of the data. From all four steps, the researchers collected 1,094 transactions of conversations between interested persons to study for the Bachelor of Information Technology via the Facebook Page of the Faculty of Information Technology.

3.3 Data preparation

The data preparation process consists of five components including selecting data, cleansing data, constructing data, integrating data, and formatting data. This research applied the principles of text mining analysis as a guideline for data preparation management. Text mining is the exploitation of messages communicated in everyday life. It is sometimes called "Text Data Mining" or "Text Analytics" [2], [30]. Text mining normally involves the process of manipulating input text structures for predictive analysis and forecasting. It performs a model acquisition function within the unstructured data into the structured data. Finally, the structured data is used for evaluating and interpreting results through a model development process for prediction and forecasting. The data preparation process is illustrated in Figure 1.

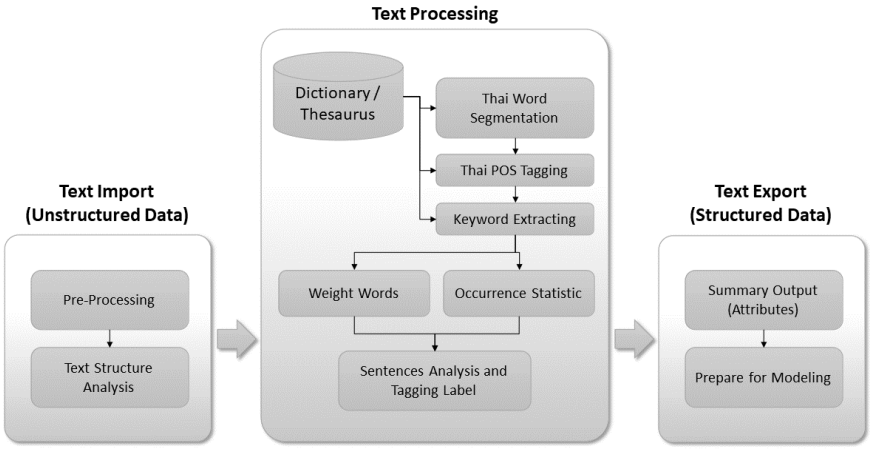

Fig. 1. The Data Preparation Process

Figure 1 demonstrates the data preparation process by applying management and analysis based on text mining principles. It has three important steps. The first step is

the text import, which is the pre-processing process. In addition, initial text structure analysis. The second step is text processing whereas the purpose of this step is to create variables for defining the model development attributes. The final step is the text export, whereby the attributes are summarized through the analysis process and transformed into a modeling-ready state.

3.4 Modeling

The CRISP-DM modeling component consists of four components [16]–[18]. The first component is the selection of modeling techniques, which considers the properties of the technique corresponding to the goals and research questions. The second component is the creation of a test design, which is used to design and select a viable technique. The third component is modeling while the last component is the assessment model.

Typically, models have competed with others in different techniques. Researchers need to interpret model results based on scientific knowledge relevant to the research problem. In addition, the researchers need to consider the criteria of success as predefined and improved test design. Therefore, the model development tools in this research were deployed by modeling machines based on supervised learning. The selected techniques consisted of three techniques including the Naïve Bayes, K-Nearest Neighbors, and Neural Network techniques. All three techniques consist of several features and details. Naïve Bayes is a classification technique that uses probability to calculate and explain equations. Moreover, the Naïve Bayes technique is suitable for data problems that are not linear model analyses and is used to rank forecasts based on probability. Therefore, it is appropriate to apply the unstructured nature (unstructured data) to create predictions that match this type of model analysis.

Likewise, K-Nearest Neighbors (K-NN) is a technique for analyzing new data compared to the original data in the immediate vicinity. If the new data is closest to the original data, it is assigned the same data type as the original data.

On the other hand, Neural Networks or Artificial Neural Network is the branch of artificial intelligence technology. It contains concepts and principles to design a computer network that mimics the work of the human brain. Although its operation is complex, it can perform analysis and prediction of data efficiently.

After defining the scope and selecting the modeling tools, the next step is to create the test design. The test design creation process is presented along with the modeling process as shown in Figure 2. Figure 2 shows the development of the model. It consists of a process based on the CRISP-DM principle in three parts. The first part is to import the prepared data from the data preparation process. The second part is performed according to the selected techniques. The third part is the selection of high-performance models where the model selection process was chosen by the cross-validation methods and confusion matrix techniques as explained in the evaluation topic.

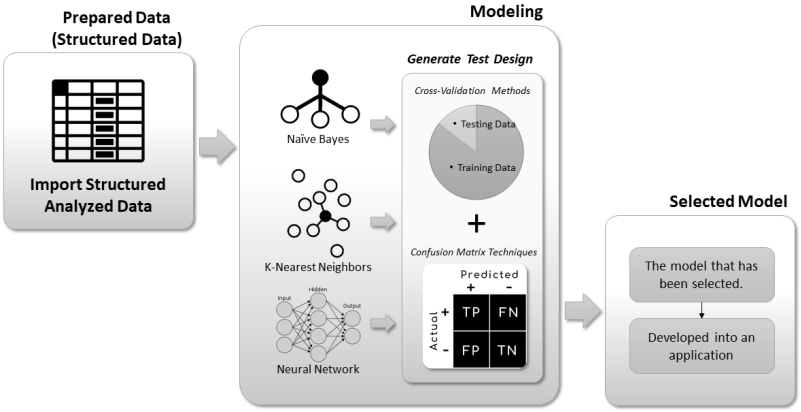

Fig. 2. Modeling and Finding the Effectiveness of Research Tools

3.5 Evaluation

The evaluation phase is an important part of considering implementing a model that has been developed [17], [30]. While the assessment model task of the modeling process focuses on evaluating technical models, the evaluation process takes a broader look at which models are most relevant to the research question and what to do next. This procedure has three tasks including evaluating the results, reviewing the process, and determining the next steps.

The techniques and methods for selecting the most efficient model consist of two parts including the cross-validation method and the confusion matrix technique [31], [32]. The principle of the cross-validation method is to divide the collected data into two parts. The first part is the data prepared to create the model. It is known as "training data". While another piece of data is prepared for testing the developed model. It is known as "testing data". The confusion matrix is a technique for determining the performance of a model using the part data prepared to test the model. It has three metric components technique including accuracy, precision, and recall. It operates in the form of a computational grid, which is described in Figure 3.

Model Accuracy : $\frac{TP+TN}{TP+TN+FP+FN}$		**Actual Class**		
		Positive	Negative	**Precision**
Predicted Class	Positive	True Positive : TP	False Positive: FP (Type 1 Error)	**Positive Predictive Value :** $\frac{TP}{TP+FP}$
	Negative	False Negative : FN (Type 2 Error)	True Negative: TN	**Negative Predictive Value :** $\frac{TN}{TN+FN}$
	Recall	**Class Accuracy Value :** $\frac{TP}{TP+FN}$	**Class Accuracy Value :** $\frac{TN}{TN+FP}$	

Fig. 3. The Composition and Calculation of the Confusion Matrix

The relationship between the research objectives and the data mining development process is shown in Table 1. After obtaining a reasonable and suitable model to be developed into an application, the researchers applied the model to Facebook chatbots as shown in the reports section. It is essential to be prepared for further satisfaction assessments.

3.6 Deployment

The fact is that the models are extremely useless unless the user can access its results and the actual implementation. Therefore, there is no doubt that good outcome models can be evaluated by real users. Therefore, the tasks of this section consist of four sub-tasks including plan deployment, plan monitoring and maintenance, producing a final report, and reviewing the project as follows.

The first sub-task was development planning for deployment, where the researchers planned to study the applied model in the Facebook chatbot program.

The second sub-task was a follow-up planning where the researchers planned to test the Facebook chatbot program in a specific sample. This sample group is 30 students from the Bachelor of Information Technology at the Faculty of Information Technology, Rajabhat Maha Sarakham University. The questionnaire was verified through the IOC process with a summary of the questions as shown in Table 2.

Table 2. Satisfaction Assessment Questionnaire for Facebook Chatbots

Stage	Assessment Issues
Stage 1: Content	
Q1	Level of satisfaction toward the information provided by the chatbot
Q2	Level of satisfaction toward the presented modernity and up-to-date information
Stage 2: Functional	
Q3	Level of satisfaction toward the interaction and language appropriateness with the chatbot
Q4	Level of satisfaction toward the accuracy in providing the information and answering the questions
Q5	Level of satisfaction toward the speed and time to provide information and answer the questions
Q7	Level of satisfaction toward the similarity with human conversation
Stage 3: Usability and Benefits	
Q8	Level of satisfaction toward the chatbot's ability to handle unexpected questions
Q9	Level of satisfaction toward the utilization and creativity of applied technology
Q10	Level of satisfaction toward the innovations which it inspired the decision to be admissions into the Bachelor of Information Technology

* Q = Question Number

The third sub-task was to test the Facebook chatbot program on a given sample as shown in Figure 4. After completion of the testing process, the tester was asked to assess the satisfaction with the questionnaire as shown in Table 2.

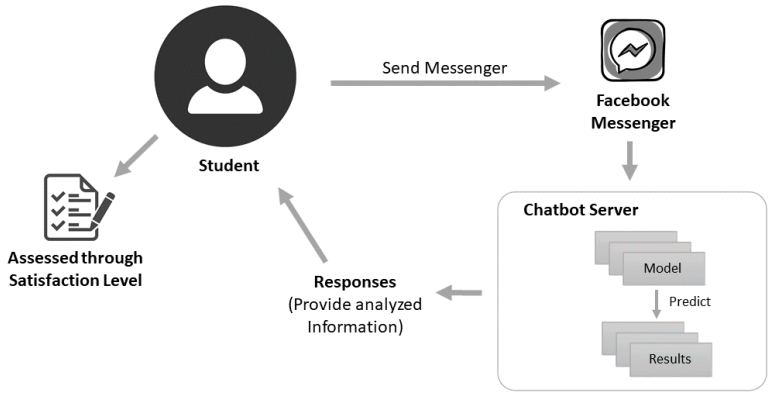

Fig. 4. Chatbot Application and Satisfaction Assessment

The final sub-task was a review of the research project where the researchers took the satisfaction assessment results from the summary tests and presented them in the research reporting section.

4 Research findings

The report of research findings was grouped into three categories by the research objectives. The first category is the text mining analytics report. This section presents a summary of the various acquired attributes. The second category is to report the analysis of the model development to be used in the Facebook chatbot program. The final category is to report on the Facebook chatbot user satisfaction assessment.

4.1 Text mining analytics report

The results of the study on the classification of questions using the text mining technique through the collection of communication transactions between interested students and the Facebook Page of the Faculty of Information Technology included 1,094 transactions from 100 Facebook users.

It discovered 55 attributes from the questions and was able to summarize the 10 most important attributes as shown in Table 3.

Table 3. The Top 10 of the Highest Frequency Attributes

Rank	Attributes	Frequency	Rank	Attributes	Frequency
1	Admission	99	6	Enrollment	77
2	Expenses	85	7	Activities	74
3	Occupation	84	8	Qualification	70
4	Salary	80	9	Registration	70
5	Scholarship	79	10	Document	70

Table 3 summarizes the filtered attributes through the text mining process, which are further used in the model analysis. These 10 attributes were designated to represent those who were discussing their interest in attending an Information Technology educational program. It is used as a "class" or "tag" for an analysis of the information provided to interested students as shown in the concept in Figure 1.

4.2 Model development report

This section reports an analysis of model development using three data mining techniques including Naïve Bayes, K-Nearest Neighbors, and Neural Network techniques. This process is illustrated by the concept in Figure 2. In addition, the testing process to determine the efficiency of each model was performed using the cross-validation technique and confusion matrix performance as described in Figure 3. A summary of the analysis results for all three techniques is shown in Table 4.

Table 4. Comparison of Accuracy using Three Techniques

Classifiers	**Model Performance**	
	Accuracy	***Time (second)***
Naïve Bayes	87.72%	45
K-Nearest Neighbors	86.56%	68
Neural Network	88.73%	145

Table 4 shows a comparison of accuracy using three techniques including Naïve Bayes, K-Nearest Neighbors, and Neural Network techniques. It found that the Neural Network technique was the most effective offering an accuracy of 88.73%. However, it took the longest time to develop the model, which was 145 seconds.

The distribution of the results of the three analytical techniques is shown in Tables 5 to Table 7.

Table 5. Summary of Naïve Bayes Technique Analysis

Class (Tag)	**Model Performance**		
	Precision	***Recall***	***F1-Score***
Admission	90.80%	85.87%	88.27%
Expenses	88.37%	78.36%	83.98%
Occupation	80.00%	81.72%	80.85%
Salary	91.25%	71.57%	80.22%
Scholarship*	72.12%	97.40%*	82.87%
Enrollment	86.25%	80.23%	83.13%
Activities*	86.76%	89.40%	92.91%*
Qualification	82.05%	76.20%	79.01%
Registration*	95.83%*	87.34%	91.39%
Document	84.88%	93.59%	89.02%

Table 5 shows the results of the Naïve Bayes efficacy test. It found that the precision that offered the greatest value across all classes was the registration class, which was 95.83%. The recall with the highest accuracy in the scholarship class was 97.40%. In addition, the F1-Score was the most valuable in the activities class, which was 92.91%.

Table 6 shows the results of the K-Nearest Neighbors efficacy test. It found that the precision that offered the greatest value across all classes was the admission class, which was 95.10%. The recall with the highest accuracy in the document class was 96.72%. In addition, F1-Score was the most valuable in the admission class, which was 90.23%.

Table 6. Summary of K-Nearest Neighbors Technique Analysis

Class (Tag)	Model Performance		
	Precision	***Recall***	***F1-Score***
Admission*	95.10%*	85.84%	90.23%*
Expenses	78.00%	92.86%	84.78%
Occupation	86.86%	91.53%	89.14%
Salary	74.16%	68.76%	71.35%
Scholarship	78.95%	87.20%	82.87%
Enrollment	82.14%	83.63%	82.88%
Activities	86.52%	90.59%	88.51%
Qualification	83.82%	67.86%	75.00%
Registration	87.14%	70.11%	77.71%
Document*	81.94%	96.72%*	88.72%

Table 7 shows the results of the Neural Network efficacy test. It found that the precision that offered the greatest value across all classes was the admission class, which was 94.44%. The recall with the highest accuracy in the activities class was 93.93%. In addition, F1-Score was the most valuable in the admission class, which was 92.39%.

Table 7. Summary of Neural Network Technique Analysis

Class (Tag)	Model Performance		
	Precision	***Recall***	***F1-Score***
Admission*	94.44%*	90.42%	92.39%*
Expenses	84.88%	87.96%	86.39%
Occupation	86.59%	86.59%	86.59%
Salary	80.73%	77.20%	78.92%
Scholarship	89.01%	75.70%	81.82%
Enrollment	82.65%	79.41%	81.00%
Activities*	84.93%	93.93%*	89.21%
Qualification	79.79%	92.60%	85.71%
Registration	76.34%	87.66%	81.61%
Document	92.86%	82.28%	87.25%

This section ends the process according to the research's second objective. The next step is to test the developed model, find satisfaction with the use of the model, and report the results.

4.3 Chatbot satisfaction assessment report

This section assesses the satisfaction of using chatbots from a model developed with a sample of 30 representatives. The assessment topics refer to Table 2, which can be summarized in Table 8.

In addition, the results of the chatbot deployment can be shown as an example of a response in the Facebook Messenger program, shown in Figure 5.

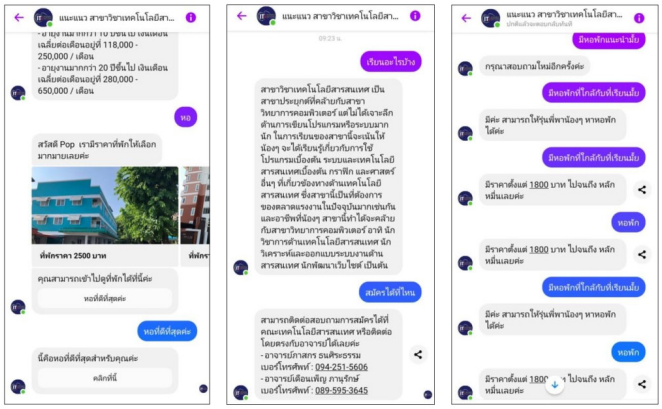

Fig. 5. An example of a response in the Facebook Messenger program

The usability assessment satisfaction score is based on the Likert 5-level evaluation principle. The definitions of the scores and their meanings are as follows: 1 means strongly disagree, 2 means disagree, 3 means neutral, 4 means agree, and 5 means strongly agree. Besides, the interpretation section is calculated from the mean and interprets the results from the score interval calculation.

$$Interval\ Scale = \frac{Maximum - Minimum}{Number\ of\ Scale} = \frac{5-1}{4} = 0.80 \quad (1)$$

The calculation of the score interval is based on the calculation in Equation (1) where each of the results is presented as follows: The first interval is the value of the mean between 1.00 - 1.80, which means "Very dissatisfied". The second interval is the value of the mean between 1.81 - 2.60, which means "Dissatisfied". The third interval is the value of the mean between 2.61 - 3.40, which means "Neither Satisfied nor Dissatisfied". The fourth interval is the value of the mean between 3.41-4.20, which means "Satisfied". The fifth interval is the value of the mean between 4.21-5.00, which means "Very Satisfied". The satisfaction assessment results for using the chatbots are summarized in Table 8.

Table 8. Summary of Satisfaction with Using Chatbots

Stage	Mean	S.D.	Interpretation
Stage 1: Content			
Q1	4.37	0.75	Very Satisfied
Q2	4.10	0.94	Satisfied
Average	4.24	0.85	Very Satisfied
Stage 2: Functional			
Q3	4.43	0.76	Very Satisfied
Q4	3.43	1.05	Satisfied
Q5	4.33	0.75	Very Satisfied
Q6	3.80	1.11	Satisfied
Q7	3.80	0.79	Satisfied
Average	3.96	0.89	Satisfied
Stage 3: Usability and Benefits			
Q8	3.93	0.81	Satisfied
Q9	3.42	1.05	Satisfied
Q10	4.09	0.94	Satisfied
Average	3.81	0.93	Satisfied
Total Average	3.97	0.90	Satisfied

Table 8 presents the results of the analysis of the satisfaction assessment of using the chatbot program. It found that the overall satisfaction with using the chatbots was "Satisfied". It has a mean of 3.97 and a standard deviation of 0.90. Moreover, most of their satisfaction was the state of the Q1: Level of satisfaction toward the information provided through the chatbot with a mean of 4.37 and a standard deviation of 0.75. Therefore, it can be concluded that the overall user accepted the developed program.

5 Research discussions

Two areas were essential to the research discussion: the model discussions and the discussion of satisfaction assessment results.

5.1 Model discussions

The models selected in the chatbot application were compared with the performance of the Naïve Bayes technique, the K-Nearest Neighbor technique, and the Neural Network technique. The comparison results of each technique are as follows:

1. Performance testing with the Naïve Bayes technique showed that the highest value Precision was 95.83% with registration class, while Recall was 97.40% with scholarship class. Also, F1-Score was 92.91% with activities class, while Accuracy was 87.72% in 45 seconds.

2. Performance testing with the K-Nearest Neighbor technique showed that the highest value Precision was 95.10% with admission class, while Recall was 96.72% with document class. Moreover, F1-Score was 90.23% with admission class, while Accuracy was 86.56% in 68 seconds.
3. Performance testing with the Neural Network technique showed that the highest value Precision was 94.44% with admission class, while Recall was 93.93% with activities class. In addition, F1-Score was 92.39% with admission class, while Accuracy was 88.73% in 145 seconds.

By selecting various research tools and carefully controlling the research process, it can be concluded that the model developed and applied is acceptable in practice.

5.2 Discussion of satisfaction assessment results

This sample group is 30 students from the Bachelor of Information Technology at the Faculty of Information Technology, Rajabhat Maha Sarakham University. The categories of the questionnaire are presented in Table 2. In addition, the level of satisfaction with the use of the chatbot program is summarized in Table 8.

It found that the overall outcome of the testers' satisfaction with the chatbots was high. It had an overall satisfaction level of 3.97, which was interpreted as Satisfied or Acceptable. The issue that the sample group had the highest level of satisfaction was the Q1: Level of satisfaction toward the information provided through the chatbot with a mean of 4.37 and a standard deviation of 0.75. This point is the main part of the research. It can be concluded that this research achieved all research objectives.

However, it has some suggestions: 1) Increasing the retention period for a more extended period will result in a more efficient analysis of user needs. 2) Problems and impacts arising from the COVID-19 pandemic may affect the decision to study in different educational institutions. Therefore, expanding the scope of research by possibly repeating the study after the end of the COVID-19 epidemic could make prediction models more accurate. 3) This research should be encouraged and supported by relevant organizations and agencies for further use in public works. All three recommendations are consistent with the research [7], [9]–[11].

6 Conclusion

In conclusion, it found that the different learners have distinctive behaviors and interests that result in manifold achievements. Presenting an educational program consistent with the learner's intentions will support the learner to achieve the desired goal. Therefore, it leads to three important research objectives: 1) to study the problem of providing information to university students with chatbots, 2) to develop a model and construct a chatbot to predict the interest of university students, and 3) to assess the satisfaction of the information provided by the chatbot application. Hence, this research achieved the objectives of the research hypothesis.

The first objective was achieved by studying the needs of learners through data collection within Messenger from the Facebook Page of the Faculty of Information Technology, Rajabhat Maha Sarakham University during the academic year 2020-2021, totaling 1,094 transactions. There were 55 attributes where the top 10 important attributes were "tag" or "class" to generate the forecasting models for providing data and information for interested parties as summarized in Table 3. The second objective was achieved by modeling and developing an application for predicting the interest of the admissions in the Bachelor of Information Technology as summarized in the model validity in Table 4. The model was accepted and developed into an application deployed from the Neural Network technique. It offered the highest accuracy with an accuracy of 88.73%. The final objective was achieved by having a high level of satisfaction from the 10 issues, shown in Table 2. Furthermore, the satisfaction assessment results were summarized in Table 8. It has an overall satisfaction rating of 3.97, and a standard deviation of 0.90.

Based on the research question, "how do educational institutions provide tools and strategies for communicating with learners to meet the need for appropriate information for the students in their educational institutions?". Researchers have developed a chatbot application for the Department of Information Technology at the Faculty of Information and Communication Technology, Rajabhat Maha Sarakham University, as shown in Figure 5, assessed by a sample of 30 representatives. They were delighted, as summarized in Table 8. Therefore, it is concluded that this research achieves the research objectives, research questions, and established research hypotheses. The results of the research are therefore determined to be put into practice. In addition, the researchers have coordinated with relevant agencies to implement this application in the next academic year at the Faculty of Information Technology, Rajabhat Maha Sarakham Maha University, for public benefit.

7 Limitations

This research has several limitations, pointing to future improvements. For researchers, finding the key points is one of the biggest challenges. Researchers have found that university students use many different platforms, and it's not just Facebook Messenger chatbots. Developing a chatbot application to cover all platforms is imperative. However, the integration of data between platforms requires further study and research in the future.

8 Acknowledgment

This research work 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, and the University of Phayao. In addition, the Faculty of Information Technology,

the Rajabhat Maha Sarakham University. The authors would like to be grateful to all of them for their support and collaboration in this study.

9 References

[1] K. Xie, G. Di Tosto, S.-B. Chen, and V. W. Vongkulluksn, "A systematic review of design and technology components of educational digital resources," *Computers & Education*, vol. 127, pp. 90–106, Dec. 2018. https://doi.org/10.1016/j.compedu.2018.08.011

[2] Q. Li, S. Li, S. Zhang, J. Hu, and J. Hu, "A Review of Text Corpus-Based Tourism Big Data Mining," *Applied Sciences*, vol. 9, no. 16, Art. no. 16, Jan. 2019. https://doi.org/10.3390/app9163300

[3] J. Rhim, M. Kwak, Y. Gong, and G. Gweon, "Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality," *Computers in Human Behavior*, vol. 126, p. 107034, Jan. 2022. https://doi.org/10.1016/j.chb.2021.107034

[4] C. W. Okonkwo and A. Ade-Ibijola, "Chatbots applications in education: A systematic review," *Computers and Education: Artificial Intelligence*, vol. 2, p. 100033, Jan. 2021. https://doi.org/10.1016/j.caeai.2021.100033

[5] N. Aoki, "An experimental study of public trust in AI chatbots in the public sector," *Government Information Quarterly*, vol. 37, no. 4, p. 101490, Oct. 2020. https://doi.org/10.1016/j.giq.2020.101490

[6] O. O. Adepoju and N. Nwulu, "Engineering Students' Innovation Competence: A Comparative Analysis of Nigeria and South Africa," *International Journal of Engineering Pedagogy (iJEP)*, vol. 10, no. 6, Art. no. 6, Dec. 2020. https://doi.org/10.3991/ijep.v10i6.14695

[7] D. Y. Mohammed, "The web-based behavior of online learning: An evaluation of different countries during the COVID-19 pandemic," *1*, vol. 2, no. 1, Art. no. 1, Mar. 2022. https://doi.org/10.25082/AMLER.2022.01.010

[8] I. Katsaris and N. Vidakis, "Adaptive e-learning systems through learning styles: A review of the literature," *1*, vol. 1, no. 2, Art. no. 2, Oct. 2021. https://doi.org/10.25082/AMLER.2021.02.007

[9] T. Karakose, R. Yirci, and S. Papadakis, "Exploring the Interrelationship between COVID-19 Phobia, Work–Family Conflict, Family–Work Conflict, and Life Satisfaction among School Administrators for Advancing Sustainable Management," *Sustainability*, vol. 13, no. 15, Art. no. 15, Jan. 2021. https://doi.org/10.3390/su13158654

[10] T. Karakose, H. Polat, and S. Papadakis, "Examining Teachers' Perspectives on School Principals' Digital Leadership Roles and Technology Capabilities during the COVID-19 Pandemic," *Sustainability*, vol. 13, no. 23, Art. no. 23, Jan. 2021. https://doi.org/10.3390/su132313448

[11] T. Karakose, R. Yirci, and S. Papadakis, "Examining the Associations between COVID-19-Related Psychological Distress, Social Media Addiction, COVID-19-Related Burnout, and Depression among School Principals and Teachers through Structural Equation Modeling," *International Journal of Environmental Research and Public Health*, vol. 19, no. 4, Art. no. 4, Jan. 2022. https://doi.org/10.3390/ijerph19041951

[12] N. H. Al-Kumaim, F. Mohammed, N. A. Gazem, Y. Fazea, A. K. Alhazmi, and O. Dakkak, "Exploring the Impact of Transformation to Fully Online Learning During COVID-19 on Malaysian University Students' Academic Life and Performance," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 15, no. 05, Art. no. 05, Mar. 2021. https://doi.org/10.3991/ijim.v15i05.20203

[13] A. Al-zubidi, N. F. AL-Bakri, R. K. Hasoun, S. H. Hashim, and H. T. S. Alrikabi, "Mobile Application to Detect Covid-19 Pandemic by Using Classification Techniques: Proposed System," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 15, no. 16, Art. no. 16, Aug. 2021. https://doi.org/10.3991/ijim.v15i16.24195

[14] R. M. Tawafak *et al.*, "Impact of Technologies During COVID-19 Pandemic for Improving Behavior Intention to Use E-learning," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 15, no. 01, Art. no. 01, Jan. 2021. https://doi.org/10.3991/ijim.v15i01.17847

[15] A. B. N. R. Putra, Y. M. Heong, D. S. Meidyanti, and A. Rahmawati, "Hi World: The Virtual Book Learning Integrated Augmented Reality to Increase Knowledge of Covid-19 Prevention in The Learning Process Post-Pandemic Era," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 16, no. 06, Art. no. 06, Mar. 2022. https://doi.org/10.3991/ijim.v16i06.29001

[16] M. Cazacu and E. Titan, "Adapting CRISP-DM for Social Sciences," *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, vol. 11, no. 2Sup1, Art. no. 2Sup1, May 2021. https://doi.org/10.18662/brain/11.2Sup1/97

[17] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Computer Science*, vol. 181, pp. 526–534, Jan. 2021. https://doi.org/10.1016/j.procs.2021.01.199

[18] 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

[19] P. Smutny and P. Schreiberova, "Chatbots for learning: A review of educational chatbots for the Facebook Messenger," *Computers & Education*, vol. 151, p. 103862, Jul. 2020. https://doi.org/10.1016/j.compedu.2020.103862

[20] C. W. Okonkwo and A. Ade-Ibijola, "Python-Bot: A Chatbot for Teaching Python Programming.," *Engineering Letters*, vol. 29, no. 1, 2020.

[21] C.-C. Liu, M.-G. Liao, C.-H. Chang, and H.-M. Lin, "An analysis of children' interaction with an AI chatbot and its impact on their interest in reading," *Computers & Education*, p. 104576, Jun. 2022. https://doi.org/10.1016/j.compedu.2022.104576

[22] O. Zahour, E. H. Benlahmar, A. Eddaoui, H. Ouchra, and O. Hourrane, "A system for educational and vocational guidance in Morocco: Chatbot E-Orientation," *Procedia Computer Science*, vol. 175, pp. 554–559, Jan. 2020. https://doi.org/10.1016/j.procs.2020.07.079

[23] F. Clarizia, F. Colace, M. Lombardi, F. Pascale, and D. Santaniello, "Chatbot: An Education Support System for Student," in *Cyberspace Safety and Security*, Cham, 2018, pp. 291–302. https://doi.org/10.1007/978-3-030-01689-0_23

[24] A. Alhabeeb and J. Rowley, "E-learning critical success factors: Comparing perspectives from academic staff and students," *Computers & Education*, vol. 127, pp. 1–12, Dec. 2018. https://doi.org/10.1016/j.compedu.2018.08.007

[25] F. Malekian and F. M. Aliabadi, "Review of Methods of Organizing the Content of the Curriculum in the Educational System, based on ICT (Information and Communication Technology) from the Experts' View," *Procedia - Social and Behavioral Sciences*, vol. 51, pp. 19–23, Jan. 2012. https://doi.org/10.1016/j.sbspro.2012.08.112

[26] S. Simões, T. Oliveira, and C. Nunes, "Influence of computers in students' academic achievement," *Heliyon*, vol. 8, no. 3, p. e09004, Mar. 2022. https://doi.org/10.1016/j.heliyon.2022.e09004

[27] J. S. Condo, E. S. M. Chan, and M. J. Kofler, "Examining the effects of ADHD symptoms and parental involvement on children's academic achievement," *Research in Developmental Disabilities*, vol. 122, p. 104156, Mar. 2022. https://doi.org/10.1016/j.ridd.2021.104156

[28] V. V. Busato, F. J. Prins, J. J. Elshout, and C. Hamaker, "Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education," *Personality and Individual Differences*, vol. 29, no. 6, pp. 1057–1068, Dec. 2000. https://doi.org/10.1016/S0191-8869(99)00253-6

[29] W. Nuankaew and P. Nuankaew, "Tolerance of Characteristics and Attributes in Developing Student's Academic Achievements," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 5, pp. 1126–1136, 2020. https://doi.org/10.25046/aj0505137

[30] C. G. Skarpathiotaki and K. E. Psannis, "Cross-Industry Process Standardization for Text Analytics," *Big Data Research*, vol. 27, p. 100274, Feb. 2022. https://doi.org/10.1016/j.bdr.2021.100274

[31] W. Nuankaew and P. Nuankaew, "Educational Engineering for Models of Academic Success in Thai Universities During the COVID-19 Pandemic: Learning Strategies for Lifelong Learning," *International Journal of Engineering Pedagogy (iJEP)*, vol. 11, no. 4, Art. no. 4, Jul. 2021. https://doi.org/10.3991/ijep.v11i4.20691

[32] P. Nuankaew and W. S. Nuankaew, "Student Performance Prediction Model for Predicting Academic Achievement of High School Students," *Student Performance Prediction Model for Predicting Academic Achievement of High School Students*, vol. 11, no. 2, pp. 949–963, Feb. 2022. https://doi.org/10.12973/eu-jer.11.2.949

10 Authors

Patchara Nasa-Ngium is currently a lecturer at the Faculty of Science and Technology, Rajabhat Maha Sarakham University, Maha Sarakham, 44000, Thailand (Email: patchara@cs.rmu.ac.th). His research interests include artificial intelligence, machine learning, evolutionary computing, and data mining.

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 of 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 2022-04-11. Resubmitted 2022-07-05. Final acceptance 2022-07-05. Final version published as submitted by the authors.