Building ESexpert: The Machine Learning System for Voice Interactive Interview Score Assessment for Student Loan Funds

Chaisiri Sanitphonklang*

Received: April 6, 2023
Revised: June 29, 2023
Accepted: August 8, 2023

* Corresponding Author: Chaisiri Sanitphonklang, E-mail: chisiri.s@chandra.ac.th
DOI: 10.14416/j.it.2023.v2.003

Abstract

Scholarship applicants' interviews generally last from an hour to an entire day. Several critical situations of each student caused the interviewer's feelings, such as stress, sadness, and anxiety. Listening to and discussing negative points can cause mental health problems. In order to decrease the duration of interviewing, This research paper discusses the development of a machine learning model for evaluating Student Loan Fund interview scores and an automated voice-based interactive machine learning interview scoring system. The process consists of building a machine-learning model using the Random Forest Algorithm with an accuracy of 99.54% and the SMOTE method to adjust the data asymmetry. The software was developed as a web application based on the SDLC for a second process, running Cloud Speech-to-Text of Google Speech API. Experts in all five areas evaluated the software, and the evaluation score was high. The research also looks at the Software Development Life Cycle (SDLC) models and the use of cross-validation and bootstrapping for accuracy estimation and model selection.

Keywords: Machine Learning, Random Forest Algorithm, Education Loan Fund, SMOTE.

1. Introduction

The Student Loan Fund aims to provide loans to students or underfunded students for tuition, education-related expenses, and necessary expenses for living during their studies. Four thousand ninety-three educational institutions participated in the project, comprising 2,993 secondary education institutions, 787 vocational institutions, and 313 higher education institutions. The scholarship application process at all institutions requires applicants to write a scholarship application that meets the requirements of the Student Loan Fund and Apply to the committee for consideration. After that, the Student Loan Fund Committee made a preliminary review of the information in the application, but there were still more applicants than for the scholarship. The next step is an in-depth analysis by interviewing applicants.

The Interview process could be more problematic for interviewers, and the interviewer must listen to the stories and concerns of the scholarship applicants. Most responses were negative and required sympathy, sadness, and depression. It can affect the interviewer's mental and emotional state and make the interviewer stressed and depressed. Mental health must therefore be looked after and prevented, thus giving rise to the twelve messages to achieve improved mental health for all and to save lives by Thomas, Steve, et al. [1]. Much more research has been done to find ways to help address mental health issues, such as Negative information measurement at ai edge by Chen M, et al [2]. In the future, artificial intelligence wisdom could provide technology that enables more compassionate and ethically sound care to diverse groups of people [3]. Nowadays, The advancement of computer science that can develop a computer system to function like a human being, known as AI [4], [5], that can think, analyze, and predict, has thus led to this research such as the prediction of epileptic seizure occurrences and the filling of a dysfunctional urinary

* *Computer Science and Artificial Intelligence, Faculty of Science, Chandrakasem Rajabhat University.*

bladder [6]. This research aims to develop a machine learning model for evaluating Student Loan Fund interview scores and an automated voice-based interactive machine learning interview scoring system.

The process consisted of 1) building a machine learning model for evaluating scores and 2) developing a scoring system in a web application format. This research paper has divided into the following sections: The first part discusses the background. The second part of this document discusses tools and algorithms. The third part is the performance report. The fourth part discusses the results. And the fifth part summarizes the results.

2. Theoretical Background and Related Researches

This research has applied various principles and theories to help increase the efficiency and reliability of the study as follows:

2.1 Machine Learning

Machine learning is the study of algorithms that learn from experience. According to the definition, Computer programs know from experience about a specific task when, as a measure of performance, a higher-performing job indicates that learning is enhanced with experience [7].

Machine learning is applied to various studies, including text and speech, sequential genetic analysis, robotics, insurance and risk, terrorist threat detection, semantic web, software engineering, computer vision, etc. Machine learning consists of several domains: learning concepts, knowledge discovery from statistical models, databases, engineering processes, and linking by mathematical algorithms [8]. The general structure for the learning process consists of a teaching aid kit Imported into an algorithm for learning. It has come out as a data model that has been tested for accuracy before putting that data model into use.

Supervised learning is one form of machine learning. The data used to teach the machine must have a label for each data set. Popular algorithms used to teach this type of data classification machine include: 1) The K-Nearest Neighbor Algorithm, or K-NN, works on reclassifying data points based on the class of most data points in K neighbor groups, where K is the number of neighbors to consider. K-NN captures the concept of similarity with a distance formula—basic mathematics. 2) Support Vector Machine, or SVM, is used for classification and regression. Its working principle is to Plot a labelled training data set on a plane and then try to find a plane that segregates data points of both colours. 3) Decision trees mimic human-level thinking. It is a tree where each node represents an attribute/feature. Each branch represents a decision, a rule, and each leaf represents an outcome. This result can be a definite value or a continuous value. They are classified in case of classification and constant in case of regression. All three algorithms were used to compare their performance against the random forest algorithm, also classified as data classification.

2.2 Random Forest Algorithm

Random Forest is a classification and regression algorithm designed for machine learning which has several advantages compared to other algorithms. It can be used for continuous and categorical datasets, dealing with outliers in training data [9]. It works with a principle similar to a decision tree and can overcome the limitations of a decision tree while preserving its benefits. Random forests multiply the number of trees and divide the features of each tree, each with its own unique features, to make each tree more diverse and independent. There are two steps for randomly.

The first step is to take the training data to create each tree at random and each tree has a different data set by the test data, called the sample data outside the bag (out-of-bag). The second step is to make predictions for each tree created in the first phase. The Working process can be explained in the below steps and diagram:

1. Select random M data points from the training set.
2. Build the decision trees associated with the selected data points.
3. Choose the number N for decision trees that you want to build.
4. Repeat Step 1 and 2.
5. For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

Jie Dou discovered that decision tree models and random forests were remarkably effective in predicting rainwater-induced landslides at Izu-Oshima, Japan, with random forests having a higher accuracy than Decision trees [10]. Similarly, Emel Sari Gokten [11] discovered the effectiveness of random forests in predicting depression and post-traumatic stress disorder in sexually abused children using a random forest classifier.

2.3 Student Loan Fund System

The Cabinet resolution established the Student Loan Fund in 1995, and this organization is a juristic person under the supervision of the Ministry of Finance. The organization aims to provide loans to underfunded students to cover tuition fees, education-related expenses, and necessary expenses for living during their studies. Four thousand ninety-three educational institutions participated in the project, comprising 2,993 secondary education institutions, 787 vocational institutions, and 313 higher education institutions.

The operation consists of four parts: 1) A student or student who is called a borrower. 2) Educational institutions. 3) Student Loan Fund and 4) the bank. The steps are as follows: First, students submit a loan application to the educational institution they are affiliated with, and The institution will then review and consider the loan application. Second, the institution offers an application to the Student Loan Fund for approval. Finally, the bank transfers the approved funds from the student loan agreement to the student. Regarding loan applications, each institution must consider selecting applicants who are short of funds and have real needs. One of the processes for consideration is to interview borrowers to obtain in-depth information to determine their suitability before submitting a loan request to the Student Loan Fund. This applicant interview process can be affected by the interviewer's mental health due to the interviewer's exposure to negative information for many hours.

2.4 Speech-to-Text

A speech-to-text person listens to what is being said and inputs it, word for word, using an electronic shorthand keyboard, speech recognition software, or a Computer Aided Transcription software system. Their keyboard or speech recognition software is linked to a computer, which converts this information to adequately spelt words [12]. Google Cloud is committed to developing artificial intelligence (AI) that is accessible to everyone and easy to control. That's why Google Cloud has developed a speech-to-text built-in user interface called Google Cloud's Speech-to-Text (STT) API [13]. The STT API lets developers convert speech into text. As advancements in AI continue to bring speech to new interfaces and devices, the STT API helps developers add speech functionality to their applications in order to better meet consumer demands.

2.5 Imbalanced Dataset

A classification data set with skewed class proportions is called imbalanced. Classes that comprise a large proportion of the data set are called majority classes. Those that make up a smaller proportion are minority classes. The imbalance was divided into three levels: 1) Mild had minority data 20-40% of the dataset, 2) Medium level had minority data 1-20%, and Extreme had minority data 1% of the dataset.

A popular solution is to augment minority data with the SMOTE technique, which Nitesh Chawla, et al described [14]. An improved SMOTE algorithm for classifying extensive asymmetric data using random forest was developed by Bhagat, Reshma C. and Sachin S. Patil. It offers better efficiency than other methods [15]. Maldonado developed a novel SMOTE that uses weighted Minkowski distances to determine the neighborhood for each minority group whose results provide the best predictive performance [16]. Equation (1) is used to generate a new sample for the SMOTE algorithm.

$$X' = X + rand(0,1) * |X - X_k| \quad (1)$$

when

instead of newly synthesized data

X instead of Original data randomly sourced

instead of Random number between 0 and 1

X_k instead of Data that are neighbors of x

2.6 System Development Life Cycle: SDLC

The software development life cycle (SDLC) is the software production process determining the direction and sequence of actions to achieve the user's desired software [17]. SDLC consists of 6 steps: 1) System Planning and understanding 2) System Analysis 3) System Design 4) System Development 5) Testing 6) Implementation. To obtain quality software and complete it on time, it is necessary to have a Module to control the production. There are various models for system development according to customer needs and project sizes, such as the Waterfall Model, Spiral Model, Iterative and Incremental Model. Adel Alshamrani and Abdullah Bahattab stated that each model is suitable for different systems. With size and time, and can also be used in combination [18].

2.7 The Cross-Industry Standard Processes: CRISP-DM

CRISP-DM [19] is a process model for data science processes. There are six steps in order: 1) Business Insight focuses on understanding the objectives and requirements of the project; 2) Data insight drives the identification, compilation and analysis of datasets that can help achieve project goals; 3) Data preparation which is often referred to as "Data Grinding" prepares the final data set for modelling, with five tasks: 3.1) Select Data, 3.2) Clean Data, 3.3) Generate Data, 3.4) Combine Data, and 3.5) Format Data: 4) Modeling. Simulation is the use of previously prepared data to teach the machine according to the learning principles of each algorithm. 5) Assessment To gain confidence in machine learning, it is necessary to evaluate its accuracy and machine reliability and 6) Deployment. Once all steps have been completed, the model is ultimately applied to solve various problems. All efforts were used for this research.

3. Research Methodology

Figure 1. illustrates the conceptual research framework, which consists of four parts: 1) Collecting data from 18 questions online from the sample group and evaluating the Committee, 2) Developing a data classification model using a random forest algorithm, and Using the SMOTE for balance data of data training set, 3) Developing the software as a web application and evaluating its performance by experts, and 4) Experts evaluate the performance of the software. It is divided into two phases as follows:

3.1 Developing an Interview Score Assessment Model with Machine Learning

Modelling process for machine learning Proceed under the procedure Cross-Industry Standard Process for Data Mining (CRISP-DM), which includes the following steps:

3.1.1 Business understanding: the objective is to create a machine-learning model for evaluating scores on the responses to 18 questions. Experts assessed the following 18 questions stating the need for grants, as shown in Table 1.

Table 1. *Interview Questions Considered by an Expert.*

No.	Question	Level of Importance
1	What are your parents' occupations?	Intermediate
2	How much is the total income of the benefactors?	high level
3	From whom is the family's income mainly derived?	high level
4	How many siblings are there in your family to take care of them?	Intermediate
5	How many brothers and sisters are working?	Intermediate
6	How is the family situation? (Cohabitation/Divorce/Separation/Died)	high level
7	What is the condition of the family's housing? (private home/welfare home/morning home)	high level
8	Does the family have any debts? If so, what is the approximate number?	high level
9	What is the type of accommodation during the study? (staying with parents/staying with relatives/renting accommodation)	Intermediate
10	Have you worked full-time or part-time before?	Intermediate

Table 1. Interview Questions Considered by an Expert.

No.	Question	Level of Importance
11	If so, what did you do before?	Intermediate
12	Are you currently working part-time?	Intermediate
13	How much do you have to pay for housing per month?	high level
14	How much do you spend on school supplies per month?	high level
15	How much does it cost to travel to study per month?	high level
16	How much do you spend on consumer goods per month?	high level
17	How much do you spend on recreational entertainment per month?	high level
18	Do you have savings? If so, how much do you collect per month?	high level

3.1.2 Data understanding : All 18 questions can be categorized into three categories.

1) The nominal scale is the data that is divided into groups.

2) Ratio data is data that has an enormous scale or measurement level.

3) Interval Scale is a data that has a spacing or the same distance can be measured, but it is a data that has no absolute zero.

3.1.3 Data preparation : Data preparation is the complete management of data, such as data cleaning, data transformation and data balance, which increases the amount of minority data or reduces the majority or mixed data so as not to cause injustices in machine learning. In this research, the data preparation steps were as follows:

1) Transform Data: Multiple-choice answers

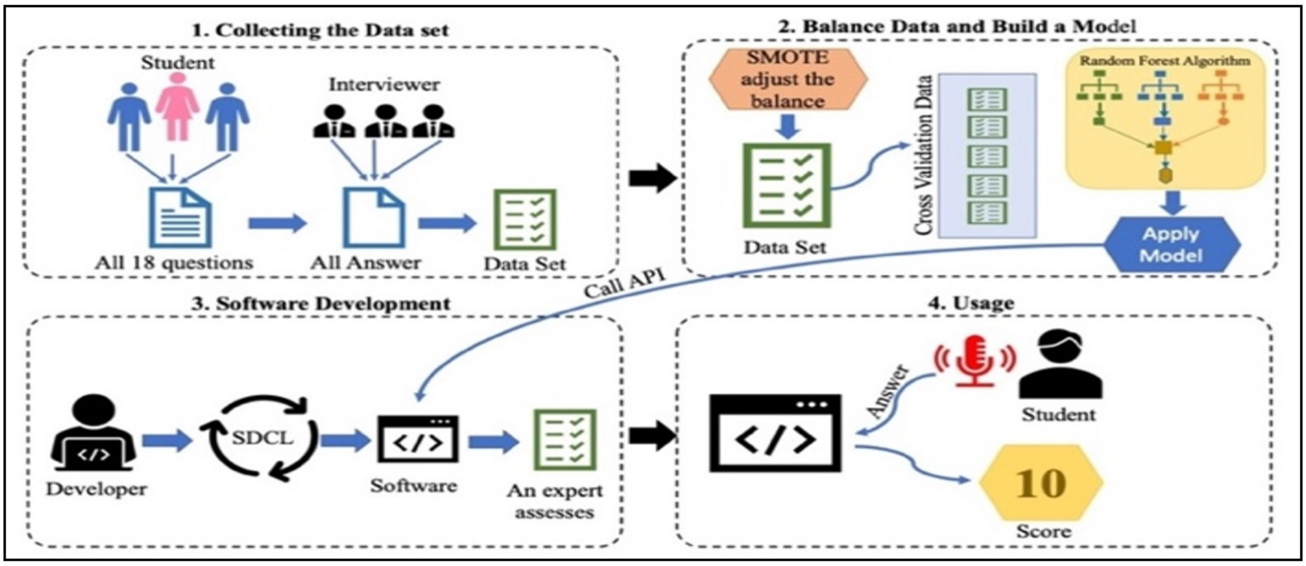

Figure 1. The conceptual research framework.

The data used in the research were the results of interviews with 1,595 students who applied for student loans fund in 2020-2021, with a selection committee of 20 people. This data was documented and collected at the Office of Student Affairs and Development, Chandrakasem Rajabhat University. Therefore, The research team has converted it into an electronic file for convenience and suitability. However, we selected only 18 questions and answers, as in Table 1 of the interviewees. characterize each question. As an example, in Table 2.

The characteristics of the completed data will be in electronic file format, Table 3.

2) Imbalanced dataset: The dataset in this research contains information on individual responses to a questionnaire for identifying the need for funding and, a second part, individual assessment scores, which are called labels. This data set contains several unbalanced labels consisting

Table 2. *Interview Questions and Answers for Student Loan Fund Scholarships.*

No.	Question	Possible Answers
1	What are your parents' occupations?	1) Employees or permanent employees 2) Business or Private affairs 3) Work for hire
2	How much is the total income of the benefactors?	1) Low income (0-10,000) 2) Medium income (10,001-50,000) 3) High income (50,001 up)
3	From whom is the family's income mainly derived?	1) Parents 2) Myself

Table 3. *Sample Data was Prepared and Imported into a Machine-Learning Model.*

Candidate No.	Answers to each question of the applicant								
	1	2	3	4	5	6	…	n	Label
1	1	1	1	2	3	3	…	3	5
2	2	1	1	1	1	3	…	3	4
3	2	1	1	1	1	2	…	2	4
4	3	2	3	1	2	1	…	1	4
5	1	3	2	1	1	1	…	1	3

of 88 Level 3, 1,452 Level 4, and 55 Level 5. The SMOT method balanced the data, which randomized new data from the original data using the principle of K-Nearest Neighbors for small data set. The SMOTE technique was used to assess the nearest neighbor value to 5, and the nominal change rate is 0.5, to add new data for the two subclasses, Level 3 and Level 5, 100, 200, 300 500 and 1,000%, respectively. The results are as follows: Level 3 contains the amount of data 176, 264, 352, 440 and 968, and for Level 5, the results are 110, 165, 220, 330 and 605, respectively. The results showed that an increasing in data by 1,000% resulted in higher prediction accuracy among minorities (accuracy = 99.54%).

Table 4. *The number of datasets Increases with SMOTE.*

Label	Increase 100%	Increase 200%	Increase 300%	Increase 500%	Increase 1000%
3	176	264	352	440	968
4	1,452	1,452	1,452	1,452	1,452
5	110	165	220	330	605
Total	1,732	1,881	2,024	2,222	3,025

A summary of the SMOTE increase data for Label 3 and Label 5 minority data is shown in Table 4. The only two classes with unbalanced data were Label 3, which accounted for 5.51%; Label 5, which accounted for 3.44% of the total data after the balance data; Label 3 had 880 data items; and Label 5, which increased 550 items.

3) Segmentation of data: The data must be divided into training and testing datasets in supervised learning. This research uses the 10-fold Cross-Validation method because every data can be tested and trained. The cross-validation method is a random sampling that starts by dividing the data into parts and examining some amounts of the data set from that data set [20]. The results of cross-validation are used as modelling options. In this case, K-fold cross-validation divides the data equally into K-series, selecting a single data to test using k-1 as training data, as illustrated in Figure 2.

3.1.4 Modelling: This experiment used a Random Forest Algorithm (RF) for data classification because it is a popular technique for classification, prediction, studying variable importance, variable selection, and outlier detection [21]. There are numerous application examples of RF in a variety of fields, such as The random forest method proves that it plays a significant role in detecting congestive heart failure (CHF) and that the random forest method gives 100% classification accuracy [22]. It was then compared to data classification with Decision Tree Algorithm (DT), Support Vector Machine Algorithm (SVM), and K-Nearest Neighbors Algorithm(K-NN).

This research conducted a machine learning process with Random Forest Algorithm by dividing the learning into three types:

Type 1 uses unsampled data supplemented for the minority

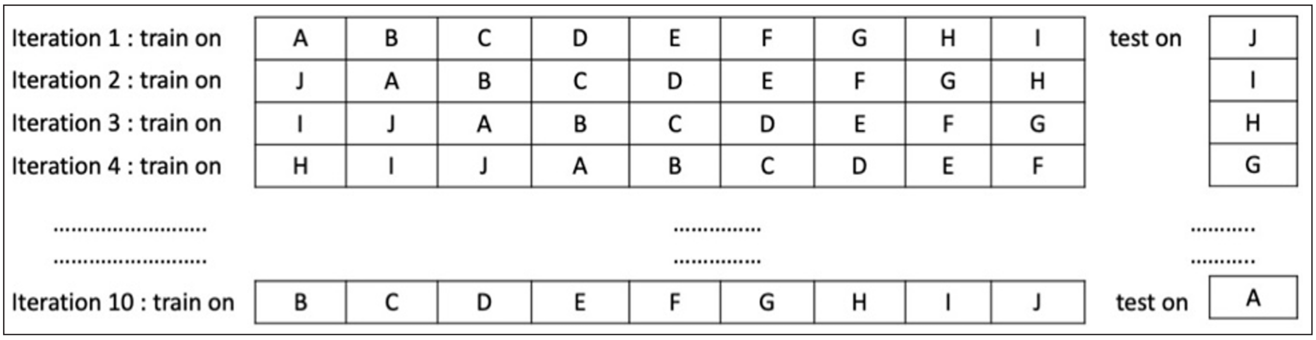

Figure 2. *10-fold Cross Validation.*

with the SMOTE method for machine learning.

Type 2 uses data with minority sample augmentation with SMOTE for machine learning.

Type 3 The second experiment was repeated with the DT Algorithm, SVM Algorithm, and K-NN Algorithm to compare the classification performance with the RF Algorithm. The results are presented in the next section.

3.1.5 Evaluation Model: Experimental results were evaluated using several statistical measures (sensitivity, specificity, F-measure accuracy and ROC curve) using a confusion matrix as illustrated in Figure 3 (for two class problems). The columns are the Predicted class, and the rows are the actual class. In the confusion matrix, TN is the number of negative examples Correctly classified (True Negatives), FP is the number of negative examples incorrectly classified as positive (False Positives), FN is the number of positive models incorrectly typed as negative (False Negatives), and TP is the number of positive models correctly classified (True Positives). Measuring the performance of the model as follows [23], [24]:

1) The Accuracy measures the model's overall forecasting performance as illustrated in Equation 2.

$$\text{Accuracy} = \frac{(\text{TPs+TNs})}{(\text{TPs+FPs+FNs+TNs})} \qquad (2)$$

2) The Recall is the probability or ratio of correct prediction in the interested group, as in Equation 3.

$$\text{Recall} = \frac{\text{TPs}}{(\text{TPs+FNs})} \qquad (3)$$

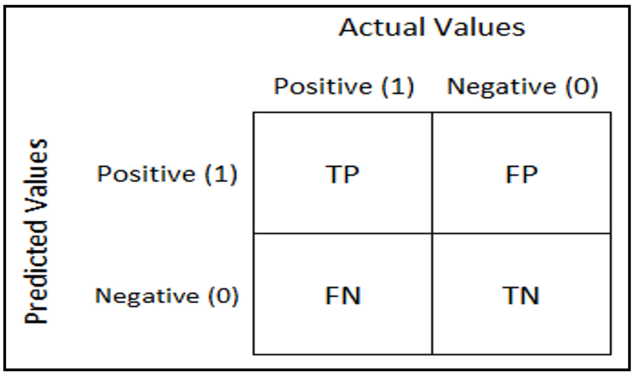

Figure 3. *Confusion Matrix.*

3) The Precision is the accuracy of predicting Positive as Positive and Predicting Negative as Negative from Equation 4.

$$\text{Precision} = \frac{\text{TPs}}{(\text{TPs+FPs})} \qquad (4)$$

4) The F1 score is a value that shows efficiency by taking precision and Recall to calculate the mean, known as the Harmonic mean. If the value is high, it indicates that the model is efficient; the calculation formula is illustrated in Equation 5.

$$\text{F1 score} = 2 \times \frac{(\text{Precision*Recall})}{(\text{Precision+Recall})} \qquad (5)$$

3.2 Software Development for Voice Interactive Interview Score Assessment of Student Loan Funds

This software development follows a six-stage software development life cycle (SDLC) within the waterfall model. If the requirement is clear for a larger project, then the “Waterfall model” was chosen that presented by S Balaji and MS Murugaiyan [25]. The framework of the interview scoring system based on voice interactions is illustrated in Figure 4. The two main areas of this framework are 1). Handling candidate information, and 2). Assessment of voice interaction scores.

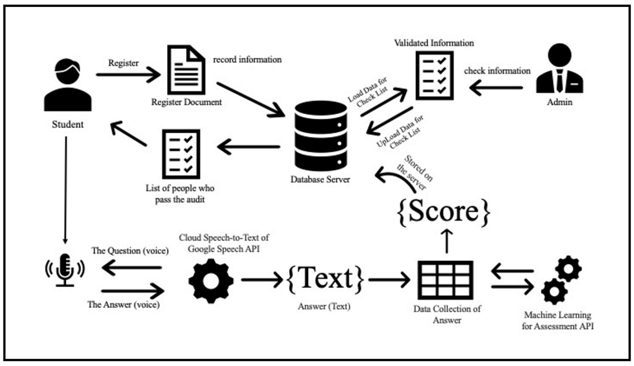

Figure 4. *The Framework of the Interview Scoring System.*

The Software development is carried out within six stages of the software development life cycle.

3.2.1 System Planning and understanding: Study and collect data from the old system by interviewing the system's users and original documents. The two main aspects of this framework are 1. Managing applicant data, and 2. Evaluating voice interaction scores.

3.2.2 System Analysis: After understanding the work system, in the next step, we analyze to divide the work into smaller parts, as illustrated in Figure 5. The system consists of 7 processes: Submit the document, Check the records, Announcement of results (1), Confirmation of admission to the interview, Get an interview, Evaluate the score, Announcement of marks (2) and Consists of 3 actors: Student, Officer and Robot.

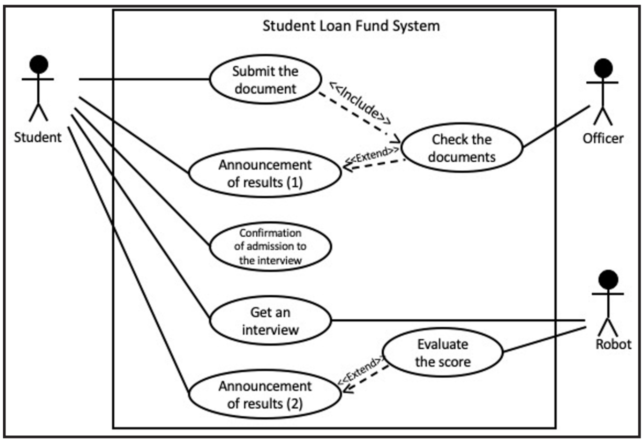

Figure 5. *Use Case Diagram of Student Loan Fund System.*

3.2.3 System Design: It is divided into three parts. The first part, Design the components of the system as illustrated in Figure 4. Students, System Administrators and Scoring Robots are considered system environments, and they had to work through the user interface was another essential part. The final component, the server, is there to store data and manage client requests. The second part, we design a database for data management. It consists of seven entities: Majors, Faculty, Student, Family, Committee and Results. These are related to each other, as illustrated in Figure 6. Finally, design a user interface that is user-friendly and easy to use, as illustrated in Figure 7.

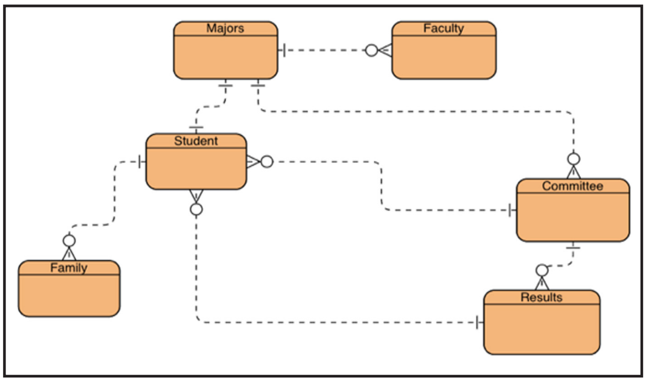

Figure 6. *ER-Diagram Diagram of Student Loan Fund System.*

Figure 7. *User Interface in of Student Loan Fund System.*

3.2.4 Our development purpose is web applications. The use of the PHP language to create functions was applied. It Contains a subscription function, submit an application function and a voice interview function. CSS, JavaScript, and HTML are used to create a form of web communication that is easy to understand for end users. The Google Speech API's Cloud Speech-to-Text converts speech to text for user answers that the robot uses to rate the response; it is implemented in the voice interview function as an API call. The result of this step is the text data converted from the speech. A machine learning API was developed in the first part of this research. The results from this process are illustrated in Figure 8.

Figure 8. *Web application of Student Loan Fund System.*

from both methods will be used to improve the program.

3.2.6 Implementation: Before deploying the system, the use of experts evaluate the software performance under ISO/IEC 25010 standards and the evaluation results are presented in the next section.

4. Result and Discussions

This research has been divided into two parts. The first is the Development of the Interview Score Assessment Model with Machine Learning, and the second is Software Development for the Voice Interactive Interview Score Assessment of Student Loan Funds. The results are reported separately as follows.

Table 5. *Machine Learning Performance after Adding Minuscule Data using SMOTE Technique with Random Forest Model.*

Label	Increase data	Performance (%)			
		Accuracy	Recall	Precision	F1
Label3	Original Data	99.19	85.23	100.00	92.03
	Increase 100%	98.45	84.66	100.00	91.69
	Increase 200%	99.15	95.83	98.06	96.93
	Increase 300%	99.21	96.59	98.84	97.70
	Increase 500%	99.23	96.63	99.10	97.85
	Increase 1000%	99.54	99.48	99.07	99.27
Label4	Original Data	99.19	100.00	99.11	99.55
	Increase 100%	98.45	100.00	98.17	99.08
	Increase 200%	99.15	99.66	99.25	99.45
	Increase 300%	99.21	99.72	99.28	99.50
	Increase 500%	99.23	99.80	99.35	99.57
	Increase 1000%	99.54	99.38	99.65	99.51
Label5	Original Data	99.19	100.00	100.00	100.00
	Increase 100%	98.45	100.00	100.00	100.00
	Increase 200%	99.15	100.00	100.00	100.00
	Increase 300%	99.21	100.00	100.00	100.00
	Increase 500%	99.23	100.00	100.00	100.00
	Increase 1000%	99.54	100.00	100.00	100.00

3.2.5 Testing: The system test has two parts: a white box and a black box. The white box is a test by a team of systems developers and computer language experts to verify the algorithms of the functions in the system. The black box is a user-tested manual to find errors through the user interface, ignoring the instructions within the program. Any errors discovered

The First: Developing an Interview Score Assessment Model with Machine Learning

This section aims to build a model to classify the data accurately; this study uses three experimental types.

1. Data preparation: There are 18 questions used in interviewing scholarship applicants, which experts have assessed

as the questions that indicate suitability for finding scholarship candidates, as shown in Table 1. Use all 18 questions for a sample of 1,595 people to answer and Subsequently, an example of 20 interview assessors to collect as a data set for machine learning. The obtained data set needs to be more balanced. In this research, the data were balanced using the SMOTE technique. The result is shown in Table 5, where adding 1,000% of minority data has the highest accuracy of 99.54% compared to adding 100%, 200%, 300% and 500% of minority data. Increase data; This column presents a onefold increase in Minority Data in each round to determine how much is enough for learning efficiency.

2. Modeling: This research uses a learning process with Random Forest Algorithm by setting the following parameters—the number of 100 random trees. Selecting the criterion on which Attributes will be selected for splitting is an information gain and set the depth of each tree to 10, which divided the learning into three types.

Type 1: Unsampled data augmented for ethnic minorities with the SMOTE method for machine learning were used for the first experiment. It was found that the classification accuracy was 99.19 %, but still classifying label 3 that is minority data wrong up to 14.77 %, as shown in Figure 9.

```
PerformanceVector:
accuracy: 99.19% +/- 0.59% (micro average: 99.18%)
ConfusionMatrix:
True:   4       5       3
4:      1452    0       13
5:      0       55      0
3:      0       0       75
```

Figure 9. *The Confusion Matrix of Type 1.*

```
PerformanceVector:
accuracy: 99.54% +/- 0.57% (micro average: 99.54%)
ConfusionMatrix:
True:   4       5       3
4:      1443    0       5
5:      0       605     0
3:      9       0       963
```

Figure 10. *The Confusion Matrix of Type 2.*

Type 2: Data were randomly added using the SMOTE method for the machine learning minority data in the second experiment. It was found that the classification accuracy was 99.54%, but the classification error for label 3 was 0.51%, which decreased from type 1 to 14.26%, as shown in Figure 10.

Type 3: The second experiment was repeated with the DT Algorithm, SVM Algorithm, and K-NN Algorithm to compare the classification performance with the RF Algorithm. It was found that the RF algorithm has a classification accuracy of 99.54%, which is the highest accuracy compared to the other three algorithms; The DT algorithm has a classification accuracy of 98.80%. We are putting the following parameters: criterion to information_gain, maximal depth to 10. The SVM algorithm has a classification accuracy of 99.36%. We are putting the following parameters: Linear Kernel Function and define Hyperparameter C as equal to 200. The K-NN algorithm has a classification accuracy of 98.45% by setting k (nearest neighbors) equal to 5 - Details are presented in Table 6. The experimental results concluded that the RF algorithm provides the highest data classification accuracy. So it is used for machine interview score estimation by creating an API to call the audio interview system.

However, the research team highly recommends the RF algorithm because it eliminates the need to determine the optimal number of neighbors, unlike the K-NN algorithm. Additionally, RF algorithms are classified as ensemble learning, producing greater confidence than the Decision Tree and SVM algorithms.

Table 6. *The Table Illustrates the Results of Comparing the Classification Accuracy Performance of the Four Algorithms.*

Algorithms	Performance (%)			
	Accuracy	Recall	Precision	F1
RF	**99.54**	**99.62**	**99.57**	**99.60**
DT	99.36	99.40	99.32	99.36
SVM	98.76	98.80	98.74	98.78
K-NN	98.45	98.50	98.42	98.47

The Second: Software Development for Voice Interactive Interview Score Assessment of Student Loan Funds.

This software consists of 3 functions, the first function is to submit documents of the applicant, and the staff verifies this document to proceed to the next step; it has the characteristics as illustrated in Figure 11.

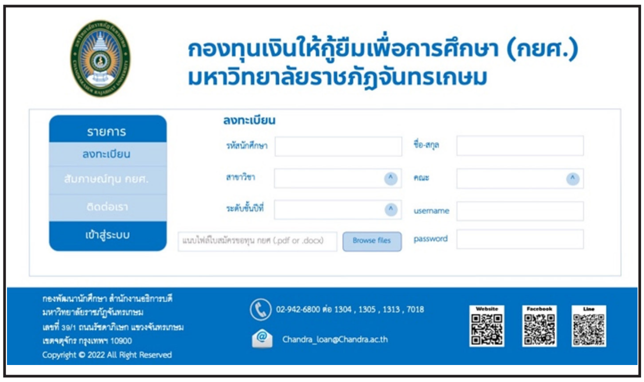

Figure 11. *Web Page for Submitting Documents.*

The second function, applicants can check the list of eligible candidates for the interview result from the documents submitted in the first step. The third function is the interview with the robotic interview system. The interview data is stored as text to lead to scoring with a machine learning model run in the API format; it has the characteristics as illustrated in Figure 12.

Figure 12. *Web Page for an Interview with the Robotics System.*

It found that the software quality evaluation results from software experts were at a high level ($\bar{x}$= 4.22, S.D. = 0.23) from the evaluation of five aspects, consisting of 1) Workability, 2) Work efficiency, 3) Difficulty - ease of use, 4) Reliability and 5) Security. Table 7 shows the results of the expert's assessment of the five aspects.

The difficulty - ease of use, it received minor marks on the evaluation for the feel of the user interface because the animations representing the interviewers lacked realism, so this area needed further development.

Table 7. *Evaluation Results of Software Quality in All Five Aspects.*

No.	Software features used in the assessment	Rating scale	
		Avg	S.D.
1	Workability	4.22	0.19
2	Work efficiency	4.44	0.19
3	Difficulty - ease of use	3.83	0.43
4	Reliability	4.13	0.12
5	Security	4.40	0.23
	Average	4.20	0.23

As can be seen from Figure 13, the bar graph presents the five performance assessment data. There were at least four areas that received high-performance ratings.

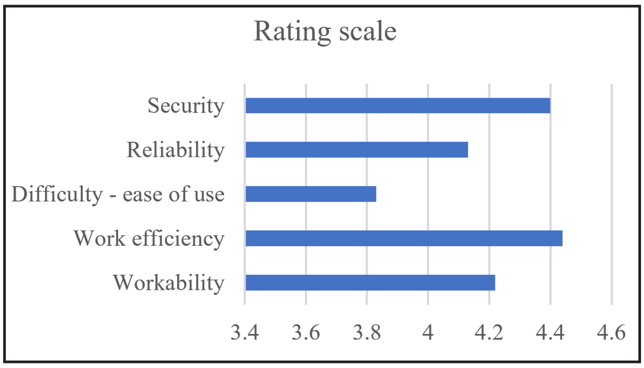

Figure 13. *The Bar Graph Presents the Five Performance Assessment Data.*

5. Conclusions

Computer technology has advanced so far that today Computers can learn to use information with machine mechanics to think analytically like humans, which we call artificial intelligence. Therefore, mental health problems caused by receiving stressful information, such as the interview committee for the selection of scholarship students who have to listen to critical situations of each student caused for many hours, can be solved or alleviated by developing an artificial intelligence system in a software platform that can replace humans with capable of human equivalents.

The process of creating this artificial intelligence consists of 2 parts:

The first part is developing a model for evaluating the applicant's response scores under Cross-Industry Standard Process for Data Mining (CRISP-DM). However, if the learning data is unbalanced, we need to balance the data with SMOTE randomization technique. If the minority data is too small than the majority data, we should randomize and Add in large enough amounts to keep the data close; for example, this time, we have increased minority data by 1000% and selected a solution-based learning algorithm that fits the data. It can be compared with other algorithms to choose the most efficient algorithm.

After that, the second part is the software implementation by developing the model from the first part to be an API and then set the software to be used.

Software must be produced within the software development life cycle to guarantee software quality according to ISO/IEC 25010. The above principles can be applied to any problem if the data is of sufficient quality and quantity for machine learning to produce accuracy in machine operations.

Another issue that should be considered in the ESexpert system is the accuracy of the voice-to-text conversion because the results impact their implementation in the evaluation section.

Future work is to develop a voice interactive system for the undergraduate admissions interview process.

6. Acknowledgement

We would like to acknowledge gratefully and thank you for the research grant support from Chadrakasem Rajabhat University Research and Development Institute (CRURDI)

7. References

[1] S. Thomas, et al. “Promoting mental health and preventing mental illness in general practice.” *London Journal of Primary Care,* Vol. 8, No. 1, pp. 3-9, February, 2016.

[2] M. Chen, K. Shen, R. Wang, Y. Miao, Y. Jiang, K. Hwang, Y. Hao, G. Tao, L. Hu, and Z. Liu. “ Negative information measurement at ai edge: A new perspective for mental health monitoring.” *ACM Transactions on Internet Technology,* Vol. 22, No. 3, pp. 1-16, January, 2022.

[3] E. E. Lee, et al. “Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom.” *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging,* Vol. 6, No. 9, pp. 856-864, September, 2021.

[4] P. H. Winston. *Artificial intelligence*. 2nd ed., Boston, United States: Addison-Wesley Longman Publishing Co., Inc., 1994.

[5] N. J. Nilsson. *The quest for artificial intelligence*. United States of America: Cambridge University Press, 2010.

[6] G. Rong, A. Mendez, E.B. Assi, B. Zhao, and M. Sawan. “Artificial intelligence in healthcare: review and prediction case studies.” *Engineering Journal,* Vol. 6, No. 3, pp. 291-301, March, 2020.

[7] T. M. Mitchell. “Does Machine learning Really Work?.” *AI Magazine,* Vol. 18, No. 3, pp. 11-20, September, 1997.

[8] M. I. Jordan and T. M. Mitchell. “ Machine learning: Trends, perspectives, and prospects.” *Science Journal,* Vol. 349, No. 6245, pp. 255-260, July, 2015.

[9] N. Horning. “Random Forests: An algorithm for image classification and generation of continuous fields data sets.” *In Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences*, Osaka, Japan, Vol. 911, pp. 1-6, 2010.

[10] J. Dou, A. P. Yunus, D. T. Bui, A. Merghadi, M. Sahana, Z. Zhu, and B. T. Pham. “Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan.” *Science of*

the total environment Journal, Vol. 662, pp. 332-346, April, 2019.

[11] E.S. Gokten and C. Uyulan. "Prediction of the development of depression and post-traumatic stress disorder in sexually abused children using a random forest classifier." *Journal of Affective Disorders,* Vol. 279, pp. 256-265, January, 2021.

[12] P. Nguyen, D. Tran, X. Huang, and D. Sharma. "Automatic classification of speaker characteristics." *In International Conference on Communications and Electronics 2010*, pp. 147-152, 2010.

[13] H. H. Wang. "Speech Recorder and Translator using Google Cloud Speech-to-Text and Translation." *Journal of Information Technology in Asia*, Vol. 9, No. 1, pp. 11-28, November, 2021.

[14] N.V. Chawia. K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." *Journal of Artificial Intelligence Research,* Vol. 16, pp. 321-357, June, 2002.

[15] R.C. Bhagat and S. S. Patil. "Enhanced SMOTE algorithm for classification of imbalanced big-data using Random Forest." *IEEE International Advance Computing Conference (IACC)*. IEEE, pp. 403-408, June, 2015.

[16] S. Maldonado, C. Vairetti, A. Fernandez, and F. Herrera. "FW-SMOTE: A feature-weighted oversampling approach for imbalanced classification." *Journal Pattern Recognition,* Vol. 124, pp. 108511, April, 2022.

[17] G.D. Everatt and R. J. McLeod. "Chapter 2: The Software Development Life Cycle." *Software Testing: Testing Across the Entire Software Development Life Cycle*. John Wiley and Sons. pp. 29–58, 2007.

[18] A. Alshamrani and A. Bahattab. "A comparison between three SDLC models waterfall model, spiral model, and Incremental/Iterative model." *International Journal of Computer Science Issues (IJCSI)*, Vol. 12, No. 1, pp.106-111, January, 2015.

[19] R. Wirth and J. Hipp. "CRISP-DM: Towards a standard process model for data mining." *In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, Vol. 1, pp. 29-39, 2000.

[20] R. Kohavi. "A study of cross-validation and bootstrap for accuracy estimation and model selection." *In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence,* Vol. 14, No. 12, pp. 1137-1143, 1995.

[21] A. B. Shaik and S. Srinivasan. "A brief survey on random forest ensembles in classification model." *In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018,* pp. 253-260, 2018.

[22] Z. Masetic and A. Subasi. "Congestive heart failure detection using random forest classifier." *Journal Computer methods and programs in biomedicine,* Vol. 130, pp. 54-64, March, 2016.

[23] J. Thongkam, G. Xu, Y. Zhang, and F. Huang. "Toward breast cancer survivability prediction models through improving training space." *Journal Expert Systems with Applications,* Vol. 36, No. 10, pp. 12200-12209, December, 2009.

[24] S.V. Stehman. "Selecting and interpreting measures of thematic classification accuracy." *Journal Remote Sensing of Environment,* Vol. 62, No. 1, pp. 77-89, 1997.

[25] S. Balaji and M.S. Murugaiyan. " Waterfall vs. V-Model vs. Agile: A comparative study on SDLC." *International Journal of Information Technology and Business Management,* Vol. 2, No. 1, pp. 26-30, June, 2012.