4th International Conference on Intelligent Engineering and Management (ICIEM 2023)

Artificial Intelligence based Chatbot for promoting Equality in High School Advising

Suha Khalil Assayed
Faculty of Engineering and Information Technology
The British University in Dubai
Dubai, UAE
sassayed@gmail.com

Manar Alkhatib
Faculty of Engineering and Information Technology
The British University in Dubai
Dubai, UAE
manar.alkhatib@buid.ac.ae

Khaled Shaalan
Faculty of Engineering and Information Technology
The British University in Dubai
Dubai, UAE
khaled.shaalan@buid.ac.ae

2023 4th International Conference on Intelligent Engineering and Management (ICIEM) | 979-8-3503-4112-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIEM59379.2023.10167112

***Abstract* -The sustainable development goal 4 (SDG4) aims to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all". Therefore, researchers are inspired to study the fairness and equality in different aspects of education. Some studies are focused on social and academic initiatives and others on developing state-of-the-art technology to enhance the education between students equally. Henceforth, in this paper a novel affordable chatbot implemented by using a neural network model and natural language processing (NLP) to assist students in high schools particularly, since high school is one of the most essential stages in students' lives, as in this stage, students have the option to select their academic streams and advanced courses that can shape their career with their passions and interests. The dataset in this study collected from different academic resources such as schools & universities websites, schools' advisers, parents and students. It includes (968) pairs of enquiries and tags. The first column represents the student's enquiry; the second column indicates the tag or the class of each sentence. The model built by connecting the input data into embedding layer, and then the data fed into the LSTM layer with different number of neurons, then authors used sigmoid function for the output layer. The result in this study shows that the performance of the chatbot is improved by increasing the number of neurons from 5 to 8, the model achieved high accuracy ratio with score (96.5%). In future the model will be developed with stacked LSTM layers with using softmax activation function in the output layer, as different classes will be added as well in the dataset.**

Keywords: chatbot, artificial intelligence, nlp, advising, high school, equality, lstm, sustainable development goals

I. INTRODUCTION

Education is an essential initiative in the 2030 agenda for sustainable development goals (SDGs); as SDG4 focused mainly on education and aims to" ensure inclusive and equitable quality education and promote lifelong learning opportunities for all". In fact, students play a significant role on improving the quality of life by using their education to solve the word's problems [1]. Therefore, researchers and authors are inspired to study the fairness and equality in education from different perspectives, Lewis et al. [2] conducted a study to investigate the initiatives that can be developed in University of Michigan in order to prepare the disadvantaged students to success academically and socially, authors provided different aspect of supports including motivation sessions, activities, as well as academic advisers and peer mentors. Moreover, Lewis et al. [2] visited different high schools which students were coming from, in order to make the problem clearer. Other studies focused on the fairness and equalities of universities admission, Assayed & Maheshwari [3] deployed a simulation agent model by using socio-economic and academic parameters such as the family-income and the high school GPA in order to maximize the utilities of the fairness and equalities of universities admission. The authors conducted this simulation in medical schools' admission particularly, since medical degree is a competitive program in most countries and the disadvantaged students with need-based would not be able to afford the cost of attending a preparation classes or having any private tutors.

A. High school Advising

High school is one of the most important stages in students' lives, as in this stage, students have the option to select their academic streams and other advanced courses that can shape their future with their passions and interests. Some students might have more interests to pursue technical and vocational education, whereas others may have different interests. Furthermore, parents are confused about the value of college degrees compared to the income afforded to graduates and the employment rate [4]. Therefore, high school advisers can influence students' decisions in selecting the right courses that fit with their skills and backgrounds, in order to prepare them to the next important milestone in post-secondary education [5].

Although college-career advisers have an essential impact on students' success and development [6], however, the disadvantaged high schools have a limited budget to hire academic advisers, some of them consider this role as a luxury and not a necessity [7]. To begin with, these underprivileged schools cannot afford hiring college-career counselors, as it is not considered a priority for them, specifically when finances are limited. As a result, students will not be properly prepared for colleges and careers compared to other students who have been subjected to advisers in their schools.

979-8-3503-4112-6/23/$31.00 ©2023 IEEE

B. Artificial Intelligence-based Chatbot

Artificial Intelligence (AI)-chatbot also known as machine-learning chatbot deals with more complex problems, which is not possible to be solved with rule-based chatbot. The most important components of the artificial intelligence that can make a successful AI chatbot are the natural language processing (NLP) and machine Learning (ML) [8], [9]. The NLP can enable the machine to understand and interpret user's request either its spoken or written, nonetheless the users' interpretation can be boosted by using the machine learning algorithms; a chatbot can learn continuously as long as it is answering more questions from users. Henceforth, in this paper a novel affordable chatbot will be implemented by using a neural network model and natural language processing (NLP) in order to understand students' questions and classify it accordingly. This paper is organized as follows: Section II explore the related works, section III describes the experiment and development, section IV describes the results and finally the conclusion and future works are described in section V.

II. RELATED WORKS

A. Chatbot in Students Advising

Despite the vital role of academic advisers in high schools but few researchers who study the dialogue systems and chatbots in this stage. Assayed et al. [10] developed a machine learning chatbot called "HSchatbot" to assist students in some of high schools to classify their enquiries based on type of their questions, however, most of the authors focused on studying the impacts of chatbots on students' admission at colleges and universities. Since most universities and colleges around the world receive thousands of applications yearly during a limited time and as a result, the admission officers would receive many enquiries from students and parents, hence having a chatbot can handle those responses effectively. Moreover, El Hefny et al. [11] developed a chatbot called "Jooka" for supporting prospective students who are targeting the German University in Cairo (GUC). Likewise, Meshram et al. [12] developed a chatbot called "college enquiry chatbot" that aims to answer any admission- related questions such as admission requirement, documents required, college fees, courses that are offered, and etc.

Therefore, in this study, the authors focused on enhancing the advising service in high schools particularly with providing an affordable and equally service to all students by deploying a deep learning model with using the embedding vectors in order to improve the accuracy in the model. Figure 1 presents the chatbot framework diagram as it shows the main components NLP and NLU, these components are responsible for understanding the meaning of the texts by using different AI algorithms and then the output is generated through the NLG by either voice or text.

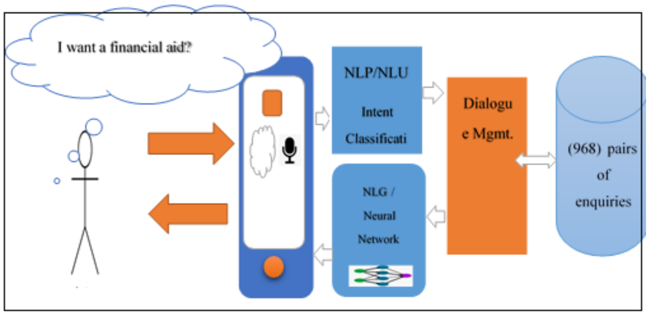

Figure 1. Chatbot framework diagram.

III. EXPERIMENT AND DEVELOPMENT

A. Corpus Collection

The dataset corpus was collected by using different academic resources, for instance universities websites, high schools' websites, students' conferences, high school's counselors, social media and students' blogs. The dataset includes (968) pairs of enquiries and labels or classes. The first column represents the question or the student's enquiry, and the second column indicates the classification label of each sentence.

B. Create a Dictionary/Vocabulary list

Building a dictionary from the corpus is an essential pre-processing step for the text in order to prepare the training corpus. It contains all unique words in the dataset and highlights the frequency of each particular word. In our model the dictionary is collected from (968) pairs of questions and labels. A split() function from Python's methods is used to split the texts into list of words. As a result, the vocab list is created as a dictionary of (1115) unique words. However, a technique of word embedding is deployed in the model in order to reduce the size of vocabulary with increasing the efficiency.

C. Data preprocessing

The purpose of this task is making the text more readable for machine learning algorithms, however, preprocessing the data includes different steps such as data preparation, cleaning the data, normalization, tokenization, padding and word embedding, in order to prepare an effective data to the machine learning which can improve the prediction accuracy [13].

D. Data annotation

Once all the dataset corpus questions and students enquires are collected, the next phase is to make sure that all the questions and enquires are classified with the correct classes by doing a manual annotation for each single row in the corpus. Annotation process is one of the main stages for building an accurate chatbot data corpus. Table 1 illustrates data classes in the dataset.

Code	Enquiries Annotation	Count
0	High school enquiry	384
1	University-inquiries	584

Table 1. The dataset classes (labels classification)

However, the class distribution in the dataset is not equal, as 60% of the data is annotated to class (1) and 40% to class (0) as shown in Figure 2.

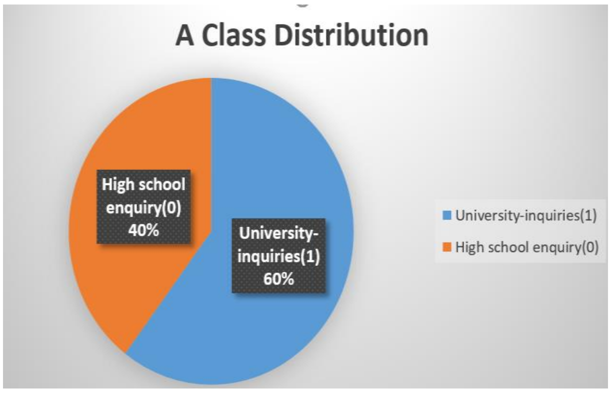

Figure 2. The distribution of the dataset classes.

E. Building the Model

In this paper, the model is implemented by using the deep recurrent neural networks (RNN) by applying long short-term memory (LSTM) network. LSTM is the advanced approach of recurrent neural network (RNN) with extending the memory. LSTM can be capable to remember inputs from long text over a long period of time by learning the order dependence in input sequences [14]. The extension memory is responsible for remembering the inputs from long sentence called "gated cell" which can give permission to this gate to open or close based on the importance of this information, however measuring the importance done by assigning weights to each part of information. Since we have a small dataset we will go for simple model by connecting the input data into embedding layer, and then we fed it into the LSTM layer, and then we used sigmoid function for the output layer. Figure 3 illustrates the structure of LSTM architecture for intent classification.

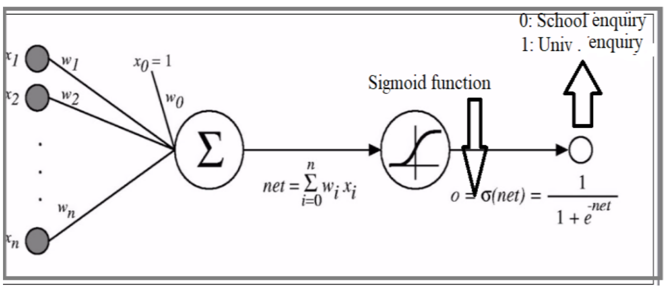

Figure 3. LSTM architecture for intent classification

In order to train the model in batches we have to be sure the length of the input sequences is the same. We used the padding function to make the length (200) for all sequences. Afterward, we fed it into the LSTM layer as shown in Figure 4.Then, the sigmoid function is added in the output layer for the binary classification, which compute the probability of the having the particular class: *School enquiries* or *Universities enquiries*.

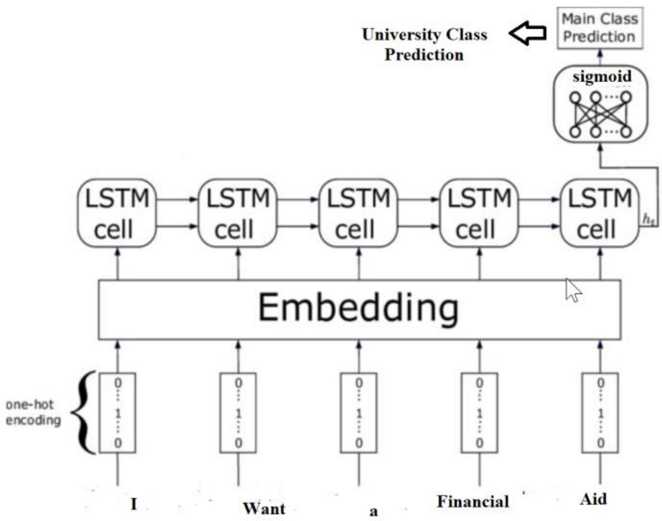

Figure 4. The LSTM architecture with predicting an example in "university class".

Results and Discussion

In this paper, the training process improved by keep tuning the hyper-parameters of the LSTM. We first selected LSTM with 5 neurons, and then increase it to the 8 neurons. The performance of the training is improved with LSTM 8, as shown in Table 2.

Table 2. The accuracy metrics with different LSTM neurons

Batch	LSTM neurons	Epochs	Accuracy
20	5	25	95.8%
20	8	25	96.5%

After that, the model is evaluated by passing the test data to the prediction function. However, 15% is randomly assigned to the test data. The below Table 3 shows the distribution of the classes in the test data. The percentage of the number of classes is almost slightly imbalanced as the ratio between them is 40:60.

Table 3. The distribution of the input data in the output classes (test data).

Code	Enquiries Annotation	Count	Ratio
0	High school enquiry	58	40%
1	University-inquiries	87	60%
Total (Test Data 15%)		145	100%

The accuracy metric is used in this model in order to calculate how often the predictions match the actual labels in the test data. Though, this metric is the most popular for the classification task, it describes the percentage of the test data that are classified correctly.
The Accuracy Metric = (Correct Pred.)/ (All Predictions)

In this model 140 out of 145 classes are matched with the target data and accordingly the accuracy score is reached 96.5 % = (140/145).

The result shows that the model is improved by increasing the number of neurons as the accuracy increased to 96.5% as shown in Table 2. Interestingly, the chatbot evaluated successfully by understanding some questions that have not trained before. As presented in Figure 5.

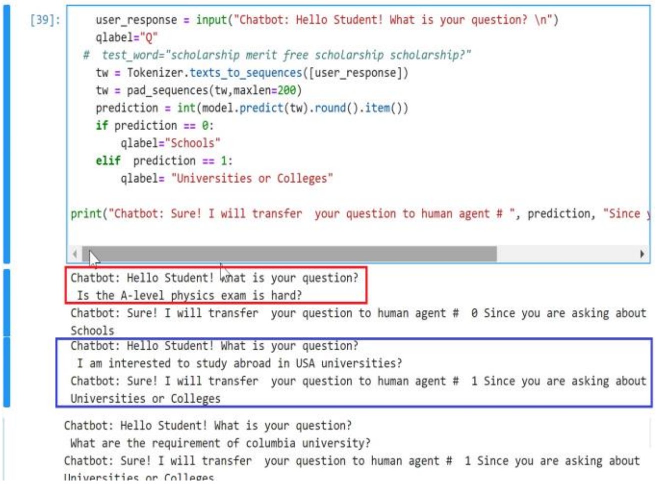

Figure 5. Testing the chatbot with new questions.

.

IV. Conclusions and Future Work

Students play a significant role on improving the quality of life by using their education to solve the word's problems Therefore, researchers and authors are inspired to study the fairness and equality in education from different perspectives, however, high school is one of the most essential stages in students' lives, as in this stage, students have the option to select their academic streams and advanced courses that can shape their career with their passions and interests. Nevertheless, the disadvantaged high schools have a limited budget for hiring advisers as they consider this role a luxury and not a necessity, these underprivileged schools cannot afford hiring college-career counselors, in this study a novel affordable chatbot implemented by using a neural network model and natural language processing (NLP) to assist students in high school advising. The model implemented by using LSTM neural network and improving the performance by tuning the hyperparameters in the learning algorithm, the accuracy ratio reached to (96.5%). In future, the model will be developed with using stack LSTM layers with softmax function in the output layer, as different classes will be added in the dataset.

References

[1] Organisation for Economic Co-operation and Development (OECD). "The future of education and skills: Education 2030." OECD Publishing (2018).

[2] Lewis Jr, N. A., & Yates, J. F. (2019). Preparing disadvantaged students for success in college: Lessons learned from the preparation initiative. Perspectives on Psychological Science, 14(1), 54-59.

[3] Assayed & Maheshwari, Agent-Based Simulation for University Students Admission: Medical Colleges in Jordan Universities (February 25, 2023). Computer Science & Engineering: An International Journal (CSEIJ), Vol 13, No 1, February 2023, Available at SSRN: http://dx.doi.org/10.2139/ssrn.4382041

[4] Curry, J. R., Milsom, A., & DEd, L. S. (2021). Career and college readiness counseling in P-12 schools. Springer Publishing Company.

[5] Haviland, S., & Robbins, S. (2021). Career and technical education as a conduit for skilled technical careers: A targeted research review and framework for future research. ETS Research Report Series, 2021(1), 1-42.

[6] American School Counselor Association. (2019a). ASCA National Model: A framework for school counseling programs. American School Counselor Association (4th ed.).

[7] Blake, M. K. (2020). Other Duties as Assigned: The Ambiguous Role of the High School Counselor. Sociology of Education, 93(4), 315–330. https://doi.org/10.1177/0038040720932563

[8] Bird, J. J., Ekárt, A., & Faria, D. R. (2021). Chatbot Interaction with Artificial Intelligence: human data augmentation with T5 and language transformer ensemble for text classification. Journal of Ambient Intelligence and Humanized Computing, 1-16.

[9] Khan, S., & Rabbani, M. R. (2021). Artificial intelligence and NLP-based chatbot for islamic banking and finance. International Journal of Information Retrieval Research (IJIRR), 11(3), 65-77.

[10] Assayed, S. K., Shaalan, K., & Alkhatib, M. (2022). A Chatbot Intent Classifier for Supporting High School Students. EAI Endorsed Transactions on Scalable Information Systems, 10(3).

[11] El Hefny, W., Mansy, Y., Abdallah, M., & Abdennadher, S. (2021, March). Jooka: A Bilingual Chatbot for University Admission. In World Conference on Information Systems and Technologies (pp. 671-681). Springer, Cham.

[12] Meshram, S., Naik, N., Megha, V. R., More, T., & Kharche, S. (2021, August). College Enquiry Chatbot using Rasa Framework. In 2021 Asian Conference on Innovation in Technology (ASIANCON) (pp. 1-8). IEEE.

[13] Misra, P., & Yadav, A. S. (2019, March). Impact of preprocessing methods on healthcare predictions. In Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE).

[14] Datta, G., Deng, H., Aviles, R., & Beerel, P. A. (2022). Towards Energy-Efficient, Low-Latency and Accurate Spiking LSTMs. arXiv preprint arXiv:2210.12613.