International Journal on Advanced Science Engineering Information Technology

Vol.12 (2022) No. 5
ISSN: 2088-5334

Intelligent Course Recommender Chatbot Using Natural Language Processing

Tan Chia Wei [a], Mohd Hanafi Ahmad Hijazi [a,*], Suraya Alias [a], Ag Asri Ag Ibrahim [a], Mohd Fairuz Iskandar Othman [b]

[a] *Faculty of Computing and Informatics, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia*
[b] *Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia*

*Corresponding author: *hanafi@ums.edu.my*

***Abstract*— Selecting elective courses in university is challenging for students as they do not know if the courses fit their interests and provide relevant knowledge or skills for their future professions. In some cases, students may register for courses without truly understanding the courses, eventually leading to course selection mistakes. A system that can recommend courses based on student's preferences is deemed necessary to address this problem. This paper proposed an intelligent course recommender system that helps students find suitable courses based on their strengths and interests. It consists of two phases. First, an intelligent course matching engine is designed and developed. The student's input is processed using natural language processing. A convolutional neural network is used to perform Part-of-Speech tagging. Keywords are identified from the processed input, and keyword matching is performed between the student's input and the courses' keywords. The most relevant courses are identified. Second, a chatbot is developed to implement the developed intelligent course matching engine. The chatbot captured student's preferences using a human-like conversation and recommended the identified most relevant courses to the students. The system is evaluated by a group of students in Universiti Malaysia Sabah. The evaluation of the usability and functionality results shows the acceptability of the proposed system, although some future work is needed based on the feedback received.**

***Keywords*— Recommender system; natural language processing; chatbot.**

Manuscript received 26 Mar. 2021; revised 14 Sep. 2021; accepted 9 Feb. 2022. Date of publication 31 Oct. 2022.

I. INTRODUCTION

Selecting available courses challenges students to make decisions based on their personal interests and talent. Some universities allow students to choose courses, such as language courses, co-curriculum courses, and most importantly, elective courses. Every course has its own purpose and scope of learning. Some courses, like elective courses, are important for the student since they may shape the knowledge or skill the student will learn. Therefore, choosing the right courses is often challenging; the selected courses need to be interesting to the students, expand the knowledge they already have, and contribute to students' professional development [1]. Choosing the right course also motivates the student and keeps the student engaged throughout their studies.

In Universiti Malaysia Sabah (UMS), students must pick the courses for each semester before the semester starts. The elements like the practicality of the course and the lecturer's teaching style are factors that affect the selection of the courses [2]. But most of the students have no clear thought on what the courses are all about, and whether the courses are suitable for them, some of them may just follow their friends [2], [3] without really knowing what the course is all about. Therefore, some of them may make mistakes in choosing suitable courses. Additionally, there is no online platform for the student from which to get course selection advice.

For an intelligent system to understand a user's query, a Natural Language Processing (NLP) pipeline is needed to analyze the query. One of the processes, POS Tagging, is challenging to apply since there are many different part-of-speech categories and variations of words. A traditional method such as rule-based POS tagging requires an expert on the language and is inefficient [4], while a statistical-based approach may produce sequences of tags that violate the grammar rule of a language [5].

In this paper, an intelligent course recommender system that is based on NLP is presented. A Convolutional Neural Network (CNN) was employed to perform the POS tagging of the text inserted by users. Keywords were then identified and matched with the existing courses' keywords. A prototype that includes a chatbot was developed, whereby the evaluation was conducted on a targeted group of students at the Faculty of Computing and Informatics (FKI), UMS.

Some background to the topics related to the proposed intelligent course recommender system is described in this section. The methodology is presented in Section 2. Section 3 presents the results and discussion. The conclusion and future work are described in Section 4.

A. Natural Language Processing

NLP is a subfield in computer science that combines both artificial intelligence and linguistics to allow the computer to understand human language [6], [7]. For the machine to understand natural language, NLP extracts the meaning by analyzing the words and the formal grammar, which specifies the relationship between words from the text. But it is challenging because the sentences used are often ungrammatical, and many ambiguous parses would be difficult to analyze [8]. Fig. 1 shows an NLP pipeline.

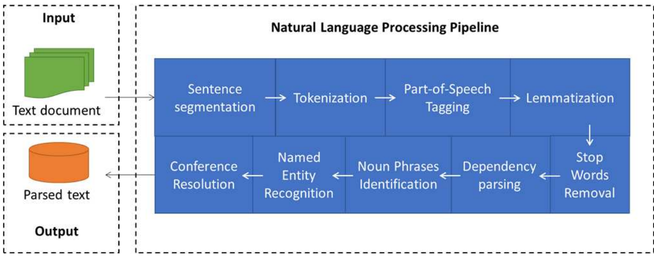

Fig. 1 NLP pipeline

An NLP pipeline consists of several processes to transform unstructured data or text into structured data. First, sentence segmentation is conducted to break apart texts into separate sentences. Second, tokenization takes place whereby sentences are separated into words. Simple tokenization can be done by splitting the word for every space or even punctuation marks. Third, POS tagging is conducted. POS tagging is an important process as it leads us to the initial understanding of what the sentences are talking about based on the role of the words. Fourth, a transformation of the inflected words into their common base form is conducted, known as lemmatization. Lemmatization decreases the ambiguity that occurs by inflectional word form. Fifth, stop words are removed from the sentences. Stop words are common words inside a text that carry less meaning. Sixth, dependency parsing is performed. Dependency parsing defines the syntactic structure of a sentence by directed grammatical relations between words [9]. Seventh, noun phrases are identified from the text. Noun phrases group words together to form meaning words that have the same meaning. Eighth, named entity recognition is conducted to classify the words into predefined categories in order to gather information from a document [10]. The predefined categories are 'who', 'where' and 'how much'. Last, a process to find all the expressions in sentences that refer to the same entity, also known as conference resolution [11], is performed. The work presented in this paper used a dialogue system, whereby the system will interact with a human in natural language.

B. Convolutional Neural Network

CNN is the most common deep learning technique used in computer vision, such as image classification and object identification, and also NLP [12]. CNN transforms image input into matrix representation and passes through multiple hidden layers, producing an output such as category or object. Fig. 2 shows an example of a CNN architecture [13].

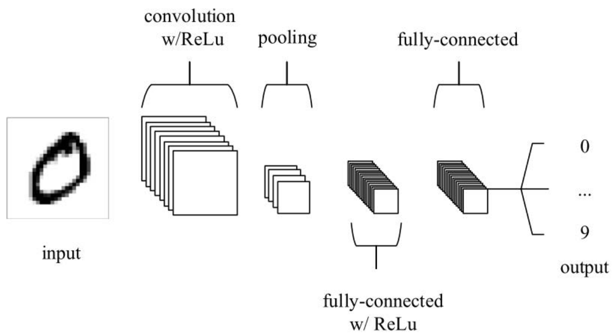

Fig. 2 An example of CNN architecture

The first layer of CNN is the convolutional layer, a 2D array representing the image's individual input. Then a filter or kernel will be created. The filter is also an array of numbers that represent the weight or parameter. The convolutional layer will be convolved with the filter by using element-wise multiplication, which multiplies the value of every respective field to produce a new layer. After the convolution is performed, a process that is known as pooling will be conducted. Pooling is a way of reducing the complexity of the network without involving any learning, and this will produce a pooling layer. The same processes are repeated until the layers are suitable to be flattened into bi-dimensional and feed into the fully connected layer.

C. Convolutional Neural Network for Text Classification

Although CNN is designed for image classification and object identification, many works employ CNN for language modeling and text classification; examples can be found in [14]–[17]. In [15], the input sentences are transformed into a word vector before feeding it into the convolutional layer. To form a convolutional layer, a convolutional filter of w □ *R*hk, where h is the window of h words covered, and k is the dimension of the word vector. For max pooling operation, the highest value cells of each layer are chosen as the representation of the other cells to put inside the fully connected layer. The output of the max pooling operation is than fed to the fully connected layer for classification. The implementation of CNN in [13] is more or less similar to the method described above with the different input sizes and repetition of the convolution and pooling process. In another work, a combination of CNN and Recurrent Neural Network (RNN) reduces the network size without compromising the classification performance [16].

D. Chatbot

A chatbot is a software that simulates natural language text or speech to interact with a human. Initially, chatbots are designed to let humans believe they are interacting with a human being [18], [19], and have been used in assisting teacher-student discussion [20]. Chatbot uses NLP and machine learning (ML), which are the subfield of AI [21]. For the chatbot to simulate natural language conversation, NLP helps the chatbot learn and understand the conversation and create natural language content [22]. ML is often used to enhance NLP performance as it helps the chatbot to learn and improve their performance through experience.

A chatbot can be an embodied or disembodied agent. Disembodied conversational agent (DCA) can usually be seen in websites, social media, and messaging apps, which generally use text-based interfaces [23]. The embodied conversational agent usually has a virtual body with nonverbal communication abilities such as body movements and facial expressions [24]. For the scope of the work presented in this paper, the type of chatbot we are interested in is DCA. More details on chatbots can be found in [18], [25].

E. Related Work

A number of works that provides academic advice to students can be found in the literature. Concerning the proposed work, two systems are reviewed in this paper due to their similarities to the proposed system: AdviseMe [26] and EASElective [27]. However, none of the work found recommended courses based on the student's preferences.

AdviseMe is a comprehensive academic advising system that provides advice for students at the Faculty of Science and Technology, University of the West Indies. The system uses the student's basic academic information such as the standing class degree, the minimum and the maximum number of credits allowed to be taken by semester, the course's prerequisites, and other structured data that can be extracted from the student's transcript. The system will provide advice through an intelligent advisor or, when necessary, a human advisor. The post-implementation phase test showed that the system could reduce human advisor workload and provide accessibility to the academic advisory to the students. The system, however, does not provide an interactive conversation platform between the students and the intelligent advisor, which is pertinent if we want to capture not only the student's historical academic achievement but also their interests and strengths.

EASElective is a chatbot that provides course information based on student's requests and captures student's opinions on certain courses through its intent detection module. To understand student's messages, NLP is employed. The evaluation of the system shows that EASElective is more convenient than the other alternatives to obtain the program's information, such as booklets and official websites. However, students require more time to understand how to interact efficiently with the chatbot.

II. Material and Method

The methodology consists of four phases, (1) requirement analysis, (2) development of the recommender engine, (3) development of the prototype of the system, and (4) usability testing (UT).

First, a specification of the course recommender system was acquired. In this phase, questionnaires were distributed to the predetermined population. The target respondents were students of the FKI, UMS. The requirements of an intelligent course recommender system are the output of this phase.

Second, the engine of the intelligent course recommender is designed and developed. Figure 3 shows the architecture of the chatbot that implements the proposed intelligent course recommendation. When a user interacts with the chatbot, the chatbot will start asking questions to get the user's reply. In this paper, the chatbot collect information concerning the user's program of study, interest, strength, experience, and future career. After the user enters the answer in natural language, the message will go through the NLP pipeline described in the preceding section. The work presented in this paper focused on the POS tagging process. A mechanism that is able to assign the correct POS tag is pertinent as it will affect the proceeding processes. This paper used CNN to perform the POS tagging and trained on the Penn Treebank Corpus version 3 dataset. CNN was selected as it is one of the best-performing and common techniques for POS tagging [28], [29]. After the message is processed, use keywords are extracted. Multiple questions are asked to identify the keywords needed as many as possible. The identified keywords from the user input are compared with the keywords of each available course using TF-IDF and cosine similarity. Further reading on how TF-IDF works can be found in Kim and Gil [30]. The courses will then be ranked based on the computed similarity score. After the relevant courses have been identified, the courses are sent to the dialogue generator. The dialogue generator uses a template-based sentence generation which maps the courses with a linguistically structured sentence. The generated sentence is shown to the user through the chatbot. The output of this phase is an intelligent chatbot that could recommend courses to students based on their preferences.

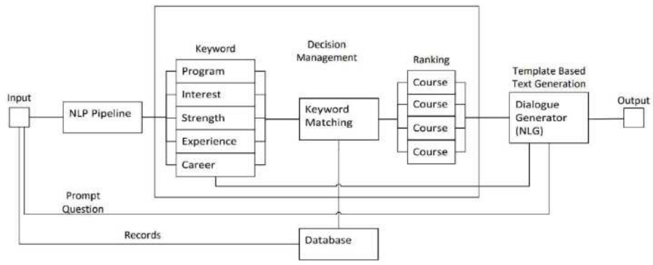

Fig. 3 The architecture of the chatbot for the proposed intelligent course recommender system

Third, the design of the course recommender system was performed. Fig. 4 shows the use case diagram of the proposed system. Two actors are Student and Admin. A student can view the faculties, programs, and course information recorded in the system. They can also get course recommendation, which is provided by the chatbot. An admin can register, update, and delete a faculty. They also can register, update, and delete programs and courses. The output of this phase is the prototype of the intelligent course recommender system.

Last, UT was conducted to obtain the user's feedback on the intelligent course recommender. Questionnaires were distributed to the target users, and their feedback on the developed intelligent course recommender system usability and functionality is captured.

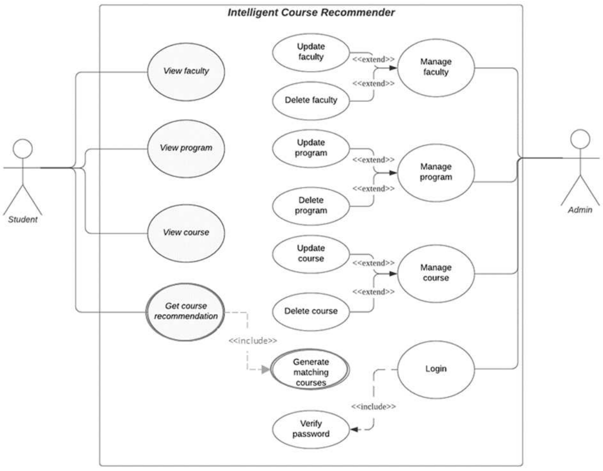

Fig. 4 Use case diagram of the proposed intelligent course recommender

III. Results and Discussion

This section presents the results obtained after the specifications gathering, the development of the proposed intelligent course recommender system, and the user feedback on the usability and functionality of the system.

A. *Requirement Analysis*

In order to identify the necessary requirement of a course recommender system, questionnaires were distributed to the target users. The target users are randomly selected amongst the third-year students of FKI, UMS. These users were chosen as they have sufficient experience in selecting courses while they were in their first and second years of study, which could provide us with the necessary feedback. A total of 35 students were identified to answer the questionnaires. Figure 5 shows the results of the questionnaire. Seven questions were asked; the answers were in the form of a Likert scale: one represents strongly disagree, and seven represents strongly agree.

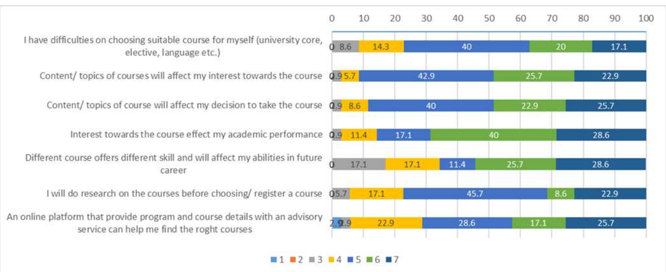

Fig. 5 Responses of the questionnaire to acquire system specification

From Fig. 5, more than 70% of the respondents agreed they faced difficulties choosing suitable courses by themselves. More than 80% also agreed that the content of the courses would affect their interest and decision whether to register for a course or not. Registering for courses of interest is pertinent as it will affect the students' performance on that particular course, according to 85.7% of the respondents. Approximately 65.7% of the respondents are also concerned that selecting the right courses will help them master the necessary skills for their future careers. Fortunately, 77.2% of the respondents admitted that they do the necessary research before registering for a course. However, currently, no platform is available that can help them choose or recommend the right courses. In conclusion, the results show that the students are in a dilemma when it comes to choosing the right courses as there are limited resources for them to do research and find suitable courses. Thus, the proposed intelligent course recommender system is needed to address this issue. Based on the feedback also, the intelligent course recommender system must have the following main features:

- Accessible online to allow easy access by the students.
- A module that will recommend courses according to the student's preferences.
- The information on the recommended courses should be made accessible.

B. *Prototype Demonstration*

In this section, the implementation of the system is presented. Fig. 6 and 7 show the main page of the system and the chatbot, whereby the system recommends courses to users based on user preferences.

Fig. 6 The main page of the intelligent course recommender system

Fig. 6 shows the home page of the proposed intelligent course recommender system. A group of icon shows the faculties available in the system. Clicking the faculty icon will send the user to the respective faculty page. The chat icon in the bottom right is the chatbot icon; a chat window will pop up when clicked, which allows the user to communicate with the chatbot. Fig. 7 shows an example of the interaction between the users and the intelligent chatbot. The chatbot asked basic information about the users, including their program of study, their existing related strength or skill, and what job types they are planning to pursue once they graduate. The recommended courses are sorted in descending order according to their similarity score to the user's preferences. The scores are calculated as described in Section 2.0. The higher the score, the more likely the course matches the users' preferences. The hyperlink to the recommended courses' page is also included to allow users quick access to the information on the recommended courses.

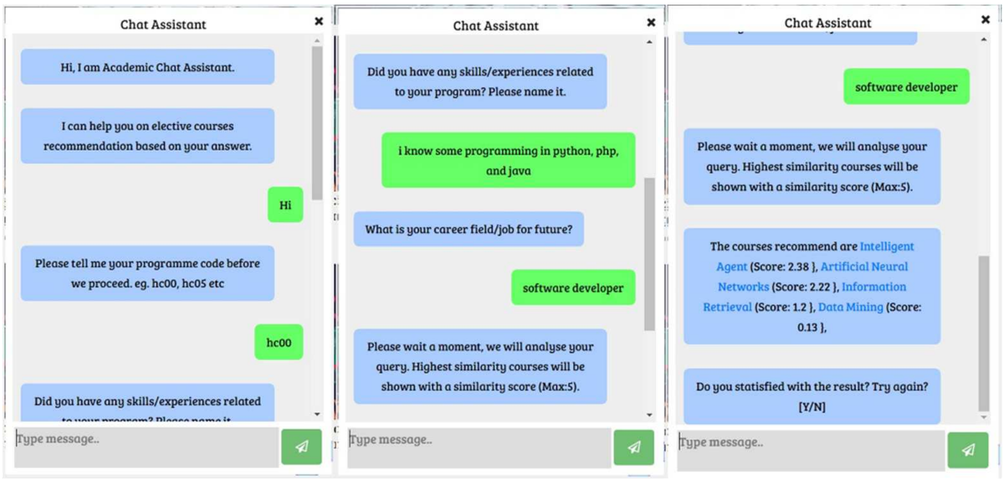

Fig. 7 Example of course recommendation based on user preference performed by the proposed system

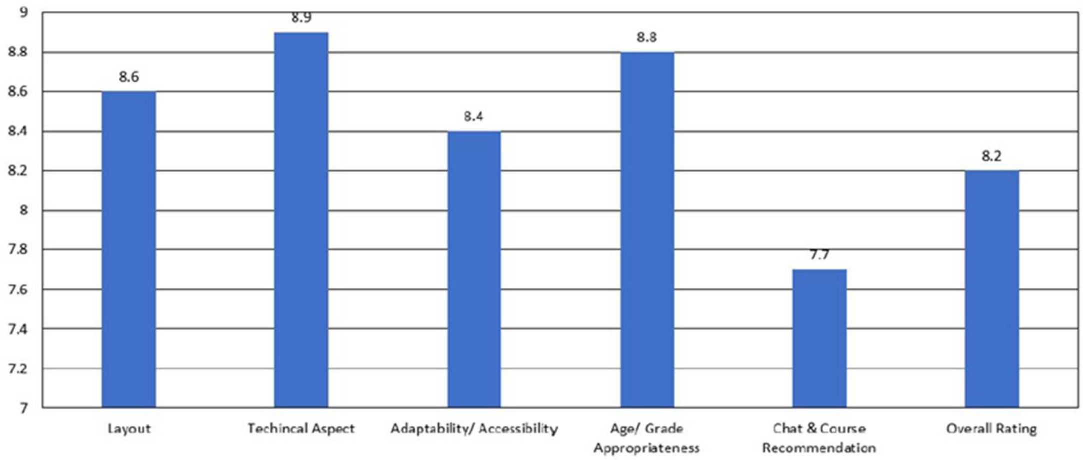

Fig. 8 Result of the UT

C. Usability Testing

For the UT, the participants were given access to the prototype, and the evaluation form was distributed to all the participants to evaluate the system. The evaluation form has been categorized into a few criteria, which include 1) layout, 2) technical aspect, 3) adaptability/accessibility, 4) age/grade appropriateness, and 5) chat and course recommendation on the student form.

Fig. 8 shows the UT rating of students towards the system using the Likert scale of 1-10; 10 being very good and 1 very bad. The figure shows that the students have a high satisfaction on most of the aspects with the highest score of 8.9 in the technical aspect, followed by 8.8 scores on age/grade appropriateness, 8.6 scores on layout, and 8.4 scores on adaptability/accessibility. Chat and course recommendations come in the lowest at 7.7 scores, which shows that the conversation or the recommendation may not appease the user. The overall acceptance score is 8.2, which indicates the usability of the proposed system is acceptable by the users.

D. Limitation and Future Work

The presented intelligent course recommender system has several limitations. First, the preferences acquired from the user are limited. This may cause some important information of the user to be not captured. Second, it is compulsory for all users to provide their responses to each question. Not all users may have an answer for each of the questions asked, and consequently, they may enter an irrelevant answer.

Based on the feedback received, some future work is identified. First, more variety of questions for course recommendations to capture more user's preferences could be incorporated into the prototype. Besides, allowing users to select which question they want to answer may improve the system's usability.

IV. Conclusion

In this paper, an intelligent course recommender system for students is presented. Each course registered into the system must be accompanied by its respective keywords. The proposed system employs NLP to understand the user's input. CNN was used to perform POS tagging due to its superiority in identifying the POS tag accurately. Keywords were then identified from the user input. TF-IDF and cosine similarity was utilized to match the user's keywords and the courses' keywords. Courses that are matched to the user's preferences will be recommended to the user. The proposed intelligent course recommender was tested in the context of students at the FKI, UMS. The UT shows the proposed system is acceptable and could be used to help students identify the right courses for them. The prototype could also be customized and used by other education providers. The code of the work presented is uploaded to the Github server and available at https://github.com/felixt8/acadadv.git. Those who are interested may access, edit and enhance the system.

References

[1] L. Movchan and I. Zarishniak, "The Role of Elective Courses in Students′ Professional Development: Foreign Experience," *Comp. Prof. Pedagog.*, vol. 7, no. 2, pp. 20–26, Jun. 2017, doi: 10.1515/rpp-2017-0018.

[2] G. Hancerliogullari Koksalmis, "Factors Affecting Selection of Elective Courses: The Use of Multi-Criteria Decision Making Model," *J. Bus. Adm. Res.*, vol. 1, no. 1, Jan. 2019, doi: 10.30564/jbar.v1i1.205.

[3] V. Roy and C. Parsad, "Efficacy of MBA: on the role of network effects in influencing the selection of elective courses," *Int. J. Educ. Manag.*, vol. 32, no. 1, pp. 84–95, Jan. 2018, doi: 10.1108/IJEM-01-2017-0005.

[4] W. Khan *et al.*, "Part of Speech Tagging in Urdu: Comparison of Machine and Deep Learning Approaches," *IEEE Access*, vol. 7, pp. 38918–38936, 2019, doi: 10.1109/ACCESS.2019.2897327.

[5] D. Kumawat and V. Jain, "POS Tagging Approaches: A Comparison," *Int. J. Comput. Appl.*, vol. 118, no. 6, pp. 32–38, May 2015, doi: 10.5120/20752-3148.

[6] D. W. Otter, J. R. Medina, and J. K. Kalita, "A Survey of the Usages of Deep Learning for Natural Language Processing," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 2, pp. 604–624, Feb. 2021, doi: 10.1109/TNNLS.2020.2979670.

[7] K. R. Chowdhary, "Natural Language Processing," in *Fundamentals of Artificial Intelligence*, New Delhi: Springer India, 2020, pp. 603–649.

[8] S. Sun, C. Luo, and J. Chen, "A review of natural language processing techniques for opinion mining systems," *Inf. Fusion*, vol. 36, pp. 10–25, Jul. 2017, doi: 10.1016/j.inffus.2016.10.004.

[9] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 2nd ed. Prentice Hall, 2008.

[10] R. E. Salah and L. Qadri binti Zakaria, "A Comparative Review of Machine Learning for Arabic Named Entity Recognition," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 2, p. 511, Apr. 2017, doi: 10.18517/ijaseit.7.2.1810.

[11] R. Sukthanker, S. Poria, E. Cambria, and R. Thirunavukarasu, "Anaphora and coreference resolution: A review," *Inf. Fusion*, vol. 59, pp. 139–162, Jul. 2020, doi: 10.1016/j.inffus.2020.01.010.

[12] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Trans. Neural Networks Learn. Syst.*, pp. 1–21, 2021, doi: 10.1109/TNNLS.2021.3084827.

[13] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," *CoRR*, vol. abs/1511.0, Nov. 2015.

[14] M. A. Parwez, M. Abulaish, and Jahiruddin, "Multi-Label Classification of Microblogging Texts Using Convolution Neural Network," *IEEE Access*, vol. 7, pp. 68678–68691, 2019, doi: 10.1109/ACCESS.2019.2919494.

[15] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1746–1751, doi: 10.3115/v1/D14-1181.

[16] A. Hassan and A. Mahmood, "Convolutional Recurrent Deep Learning Model for Sentence Classification," *IEEE Access*, vol. 6, pp. 13949–13957, 2018, doi: 10.1109/ACCESS.2018.2814818.

[17] Z. Wang, S. Cui, X. Kang, W. Sun, and Z. Li, "Densely Connected Convolutional Network for Audio Spoofing Detection," in *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2020, pp. 1352–1360.

[18] M. Skjuve, A. Følstad, K. I. Fostervold, and P. B. Brandtzaeg, "My Chatbot Companion - a Study of Human-Chatbot Relationships," *Int. J. Hum. Comput. Stud.*, vol. 149, p. 102601, May 2021, doi: 10.1016/j.ijhcs.2021.102601.

[19] J. Cahn, "CHATBOT: Architecture, Design, & Development," University of Pennsylvania, 2017.

[20] D. Feng, E. Shaw, J. Kim, and E. Hovy, "An intelligent discussion-bot for answering student queries in threaded discussions," in *Proceedings of the 11th international conference on Intelligent user interfaces - IUI '06*, 2006, p. 171, doi: 10.1145/1111449.1111488.

[21] R. Pereira, M. J., Coheur, L., Fialho, P., & Ribeiro, "Chatbots' Greetings to Human-Computer Communication," in *CEUR Workshop Proceedings*, 2018, pp. 61–66.

[22] S. Ayanouz, B. A. Abdelhakim, and M. Benhmed, "A Smart Chatbot Architecture based NLP and Machine Learning for Health Care Assistance," in *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*, 2020, pp. 1–6, doi: 10.1145/3386723.3387897.

[23] T. Araujo, "Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions," *Comput. Human Behav.*, vol. 85, pp. 183–189, Aug. 2018, doi: 10.1016/j.chb.2018.03.051.

[24] S. Provoost, H. M. Lau, J. Ruwaard, and H. Riper, "Embodied Conversational Agents in Clinical Psychology: A Scoping Review," *J. Med. Internet Res.*, vol. 19, no. 5, p. e151, May 2017, doi: 10.2196/jmir.6553.

[25] W. Villegas-Ch, A. Arias-Navarrete, and X. Palacios-Pacheco, "Proposal of an Architecture for the Integration of a Chatbot with Artificial Intelligence in a Smart Campus for the Improvement of Learning," *Sustainability*, vol. 12, no. 4, p. 1500, Feb. 2020, doi: 10.3390/su12041500.

[26] L. Keston and W. Goodridge, "AdviseMe: An Intelligent Web-Based Application for Academic Advising," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 8, 2015, doi: 10.14569/IJACSA.2015.060831.

[27] C. Chun Ho, H. L. Lee, W. K. Lo, and K. F. A. Lui, "Developing a Chatbot for College Student Programme Advisement," in *2018 International Symposium on Educational Technology (ISET)*, 2018, pp. 52–56, doi: 10.1109/ISET.2018.00021.

[28] S. Meftah, N. Semmar, and F. Sadat, "A Neural Network Model for Part-Of-Speech Tagging of Social Media Texts," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018, pp. 2821–2828.

[29] X. Yu, A. Falenska, and N. T. Vu, "A General-Purpose Tagger with Convolutional Neural Networks," in *Proceedings of the First Workshop on Subword and Character Level Models in NLP*, 2017, pp. 124–129, doi: 10.18653/v1/W17-4118.

[30] S.-W. Kim and J.-M. Gil, "Research paper classification systems based on TF-IDF and LDA schemes," *Human-centric Comput. Inf. Sci.*, vol. 9, no. 1, p. 30, Dec. 2019, doi: 10.1186/s13673-019-0192-7.