

Using Recommender Systems for Matching Students with Suitable Specialization: An Exploratory Study at King Abdulaziz University

<https://doi.org/10.3991/ijet.v16i03.17829>

Khloud Alshaikh ^(✉), Naela Baharmuz, Ola Torabah, Sara Alzahrani,
Zainab Alshingiti, Maram Meccawy
King Abdulaziz University, Jeddah, KSA
khassanalshaikh@stu.kau.edu.sa

Abstract—In Saudi Arabia, all high school graduates who want join local universities have to go through a preparatory year before selecting their specific specialization/major. One of the most concerning issues for those fresh undergraduate college students is the selection of their specialization. College specialization selection is critical for them, as their academic and career future will be affected by this decision. An unsuitable specialization selection will have unfortunate consequences, not only on the students' future but also on the university's resources and budget. This paper suggests a solution to this problem by introducing a preliminary study of a recommender system (RS), which will recommend the appropriate specialization for the students based on various tests and grades during the preparatory year at King Abdulaziz University (KAU). The proposed system guides students through their specialization selection process based on their abilities. The collaborative filtering technique was used to build the RS and K-fold cross-validation was adopted to evaluate its accuracy and performance. The results showed the prediction of a specialization for each student with good accuracy ratio. These promising initial results provide a feasible solution to assess this issue further in future studies.

Keywords—Artificial Intelligence, Education, Machine learning, Recommender Systems

1 Introduction

Choosing the right specialization for a fresh college student is one of the most critical decisions in his/her future life. Due to the young age of the fresh college student, they generally might lack the experience to make such a significant decision. However, many factors affect the choice of an appropriate specialization [1]. The best-known factor that leads students to fail to select a suitable specialization is their lack of deep knowledge about the specialization. The compulsory preparatory year in Saudi Arabian universities only enlightens the student with a broad overview of the essential subjects in different specializations. Some students may prefer to decide on a specialization by using a quick Web search but this is not enough to grasp different aspects of

it. Others may select their specializations based on a friend's recommendation, or because it is currently popular among their peers. In addition, parents and social pressure could have a huge impact on their decisions. Their family's socioeconomic background can also affect this decision [2]. An unsuitable specialization for fresh students has serious consequences for students and universities, with low grades related to higher switching rates, but these specialization switches do not lead to grade improvement [3]. Selecting the appropriate specialization is a key determinant of future earnings and career progression [1]. Accordingly, selecting inappropriate specialization can lead to unsuitable job and low earnings [1, 4]. Furthermore, inappropriate specialization is a waste of the university's resources and budget [1]. To solve the problem, this research introduces a Recommender system (RS), which in this context is an artificial intelligence technological assistant tool to support and guide user decisions with relevant information. This can be a great solution for fresh college students, as it efficiently decides the best fit specialization, based on students' preference and information [5].

This paper introduces a preliminary study that aims to explore the importance of a RS for fresh college students' specialization at King Abdulaziz University (KAU). It proposes KAURS to help students select the most appropriate specialization by the end of the preparatory year.

2 Literature Review

Recommendation system (RS) are software or tools providing recommendations to the user of items they might be interested in [6]. RS are used across many sectors to pave the way for users to make better choices online. It could be used to recommend a certain product, movie, or an online course. For example, work presented in [7] proposes a recommender system to promote employability by showing users their skills gap and match it with the right course. Another example of using RS in the educational field is presented in [8], where a hybrid model for an intelligent recommender system for e-learning platforms using data mining is introduced. Moreover, an intelligent classroom model with adaptive learning resource recommendation was proposed in [9] to cater for individual characteristics of the students, where results showed that such recommendations could improve students' learning efficiency.

Recommender system has many powerful techniques such as: decision tree, content-based filtering, and collaborative filtering. A few studies in the RS field depend solely on Decision tree approach, most combine the decision tree approach with another RS approach to achieve greater accuracy, such as the association rule approach [6]. In the study [10], decision tree and association rule approaches were utilized to build an accurate institutional RS, which aims to help students select appropriate universities based on their context and educational institution information in a mobile environment. The study's evaluation test used the cross-validation method for the two approaches, and achieved an accuracy level of 69.03% for suggesting the most suitable university. A university in Thailand [11] built a RS to help students select suitable majors based on courses and registration information. A combination of the decision

tree and Neural Network approaches has been applied in different studies. For example, a new RS was built to predict specialization and study track for students based on grades [12]. Adding the Neural Network and decision tree provides better results with a small accuracy difference. A case study was conducted in the study [13] by Delhi Technological University. Another RS was developed using a Neural Network with the decision tree technique, and the Weka and SPSS Clementine Predictive tools to determine the accuracy level in the predictive analytics process that predicts enrolment decisions, future grades, and satisfaction level. The accuracy level was 99.95%. However, the study proposes a different approach [14], the R-Tree algorithm, a multi-dimensional indexing tree data structure that is very useful for the efficient management of huge training datasets [14].

Content-based filtering is usually used in education, where many studies are conducted. One of these is mentioned in [15]. This proposed a RS that predicted the student's course selection based on marks and job interest, with the clustering technique. The study [16] proposed a course RS to improve the students' future career, built using content-based filtering techniques and an ensemble learning algorithm with k-means clustering (k number of centroids) and Term Frequency-Inverse Document Frequency (TF-IDF) keyword extraction [16]. It proved the ensemble approach is more accurate for long query terms with multiple terms, but traditional keyword extraction techniques are more suitable for short query terms [16].

Collaborative filtering is often used in education, for instance, study [17] proposed a RS that predicted elective subjects for new students by comparing and analyzing personal records and past behavior. Furthermore, [18] the proposed system recommended elective courses based on affinity between courses taken by similar students. Two studies proposed a system recommending elective courses. Neural Network and association rules were applied in [17], but in study [18] only the association rule was applied. Otherwise [19], researchers used the correlation threshold and nearest neighbor approach. [20] proposed a system recommending optional courses. Former students' enrollment records were examined to detect similarity and predict optional courses using the Pearson Correlation Coefficient and Alternating Least Square (ALS) algorithm [20]. In addition, a RS for Spanish schools applied the Collaborative technique [21].

This literature review revealed the most common three techniques used in RS designed to guide students. The Collaborative filtering proving accurate in some studies [20]. This paper adopts the collaborative filtering techniques to develop an accurate RS for KAU students. The specialization recommendation system has never been applied to KAU before. Hence, there is a need for a RS that helps KAU students select the most appropriate specialization. This paper introduces a preliminary study that explores the requirements of an accurate and reliable specialization RS for KAU students. It attempts to answer the following research questions:

- How can we minimize the possibility of drop-off or failure between KAU preparatory year students by using the KAURS system that guides them to choose the suitable specialization?
- What is the most plausible algorithm to build an accurate KAURS system that can assist KAU preparatory year students to select their specialization?

3 Data Gathering

Before implementation, appropriate and consistent data are needed so the system's results are acceptable. KAU employees were interviewed to gather data for the system and collect enrolment records from the university registrations' database. Interviews were conducted with employees from the *Deanship of Admission and Registration* and *Deanship of Information Technology* about the admission system and to request student data from the KAU database. This research focused on recommending four specializations in KAU, Science, Medical, Computer, and Engineering to narrow the related research field for obtaining accurate results. Data about each college's specialization and enrolment requirements came from the university website. The system dealt with three groups of required data, collected from the student record on the KAU database shown in Table 1. The first group consists of two elements, the exams by the National Centre of Assessment [22]. The second consists of ten elements, the preparatory year course exam results. The third group consists of two elements, the Student Cumulative Grade Point Average for two levels. The real dataset has information on 960 KAU preparatory year students in enrolment year (2017).

Table 1. Groups of data required for Implementation

Group (1)	Group (2)	Group (3)
General Aptitude Test (GAT) ¹	English Course Level (102) (ELI 102)	Student Cumulative Grade Point Average (GPA) in Preparatory year. (P_GPA)
	English Course Level (103) (ELI 103)	
	English Course Level (104) (ELI 104)	
	Mathematics Course (MATH)	
	Physics Course (PHYS)	
Scholastic Achievement Admission Test (SAT) ²	Biochemistry Course (BIO)	Student Cumulative Grade Point Average (GPA) in the current level. (C_GPA)
	Chemistry Course (CHEM)	
	Communication Skills course (COMM)	
	Computer Skills course (CPIT)	
	Static Course (STAT)	

¹ Measures the analytical and deductive skills of the student, it focuses on testing the general ability for learning regardless of any specific skills in a certain subject or topic.

² Covers the general and key concepts in biology, chemistry, physics, mathematics and English covered in the courses of the three grades of General Secondary School.

4 The Methodology

The RS is based on a collaborative filtering algorithm, which predicts certain factors for the target user, based on the best matching results from users with a similar experiment or preference [23]. Collaborative filtering algorithms contain two steps. First, find the similarity between two students by using the Pearson correlation coefficient algorithm, where similarity is based on the marks of different factors, as mentioned in section 3. Second, use the k-nearest-neighbours algorithm to predict the specialization, which relies on like-minded people or other trusted sources to predict the value of an item [6]. The two steps will be detailed next.

4.1 Similarity

To predict a student's specialization, k-nearest-neighbours should be found first (k-nearest-neighbours is a set of students with the highest similarity to any student based on their marks). To discover the similarities between students, the Pearson correlation coefficient algorithm is used, based on the marks of different subjects. The formula of this coefficient is given in equation 1 [23]. To calculate similarity W_{uv} between two students u and v , the algorithm used three parameters:

1. $i \in I$ Summation of the items that both students u and v have grade more than zero.
2. \bar{r}_u are the average marks of the u user of subjects that both students u and v have grade more than zero.
3. \bar{r}_v are the average marks of the v user of subjects that both students u and v have grade more than zero.

$$W_{u,v} = \frac{\sum_{i \in I} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I} (r_{v,i} - \bar{r}_v)^2}} \quad (1)$$

4.2 Prediction

The nearest-neighbours algorithm is used in the second step to predict the right specialization for a student. The nearest-neighbours algorithm formula is given in equation 2 [6]. Students in the datasets has different specialization. Thus, the nearest-neighbors algorithm is used multiple times, to predict rate r_{ui} for each specialization. And the highest r_{ui} predicted is a student specialization. To predict r_{ui} , a rating of specialization i for a student u , the algorithm used three parameters:

1. $v \in Ni$, where v is each student in k-nearest-neighbors set Ni
2. r_{vi} , which are marks of v in each subject i .
3. W_{uv} , representing the similarity between v and u , which has been calculated in section 2.1.

$$\check{r}_{ui} = \frac{\sum_{v \in Ni} W_{uv} r_{vi}}{\sum_{v \in Ni(u)} |W_{uv}|} \quad (2)$$

5 System Analysis and Design

KAURS system uses real data from the King Abdulaziz University database, as shown in Figure 1. First, the student logs in using his ID and password provided by KAU through the interface. Second, the student's information is retrieved from databases and implemented by the recommendation engine, which applies collaborative filtering to generate suitable specializations.

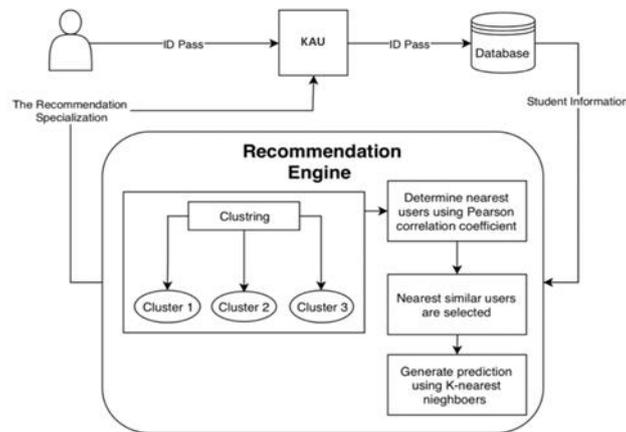


Fig. 1. System Architecture Design

6 Experiment and Results

The system has a dataset containing 960 records of science, medical, computer, and engineering students to train the system. Each record has marks of (GAT, SAT, ELI 102, ELI103, ELI104, BIO, CHEM, COMM, CPIT, STAT, MATH, PHYS, P_GPA, C_GPA), as shown in Table 1. The system starts when a student enters his/her ID to fetch his/her data from the test dataset. The Pearson correlation coefficient algorithm is used to calculate the similarity percentage between the student and all students' records in the learning dataset to find the 3 k-nearest-neighbours. To predict the student specialization, the system first checks C_GPA in the three k-nearest-neighbours. If all 3 k-nearest-neighbours C_GPA is lower than 4.5, the system prints "your marks are not good enough for the four specializations". If not, the user-based neighbourhood algorithm is applied to 3 k-nearest-neighbours to find the specialization.

To evaluate the accuracy and performance of the k-nearest-neighbours algorithm, we use two methods. The first method, the datasets split into two datasets: 20% of the dataset for testing and the rest 80% for training, it generates 70.83% accuracy. The second is the k-fold cross-validation method was applied. This splits the learning dataset to K smaller sets, then applies the test K times. In each test, one set is used for test and the others for training [24]. This is computationally expensive but beneficial to the KAURS system as the dataset is not big and allows the evaluating phase to use

larger sets for training and testing. In evaluating this method, it has been applied twice were the datasets split 5 times and 10 times. Table 2 shows the cross-validation results. Thus, the KAURS system’s highest accuracy is 74.79%. The result was sufficient to predict a specialization and give a good accuracy ratio that gives a feasible solution to be assessed further in future studies.

Table 2. The results of k-nearest-neighbors algorithm

Classifier	20/80	Five cross-validation	Ten cross-validation
k-nearest-neighbours	70.83	73.23	74.79

7 Discussion, Conclusion and Future Work

After recapping the previous studies in the recommendation system. This paper proposed to build the KAURS system which has never been applied to KAU before. It was proposed after considering experiences of solutions in the RS that could be applied in the educational field and comparing them to similar studies in the literature. The KAURS system results are compatible with those previous results. The initial results of testing KAURS system was sufficient and showed that there is need for it in KAU in Saudi Arabia.

As this is a preliminary study, the dataset was minimized to cover four specializations only. Larger datasets will be used to draw a more general conclusion.

Selecting suitable specialization is a critical decision for all preparatory year students in Saudi universities. Therefore, this paper proposed the KAURS system to guide the fresh student in selecting the most appropriate specialization based on their scores. The RS developed for Science, Medical, Computer and Engineering colleges students uses a collaborative filtering technique based on real students' data. The experiment result showed that the recommendation was sufficient to predict the specialization with a good accuracy ratio.

The future work of this paper will be applicable at all specializations in KAU by integrating within the official KAU Website to guide students to choose their educational path. science, medical, computer, and engineering

8 Acknowledgement

The authors want to express their gratitude to the Deanship of Information Technology at King Abdulaziz University for supplying the dataset used in this research. In addition, the authors wish to thank Deanship of Admission and Registration for the Information provided to complete this work.

9 References

- [1] A. A. Alkhail. (2017). Professional self-efficacy and its relationship to the difficulty of making professional de-cision among a sample of preparatory year students at King

- Abdulaziz University - Jed-dah (no. 002). . Available: <http://search.shamaa.org/FullRecord?ID=124377>
- [2] W. Chen, "Linear Networks and Systems (Referencia de libro)," Belmont, CA: Wadsworth, pp. 123-135, 1993.
- [3] C. Astorne-Figari and J. D. Speer, "Are changes of major major changes? The roles of grades, gender, and preferences in college major switching," *Economics of Education Review*, vol. 70, pp. 75-93, 2019. <https://doi.org/10.1016/j.econedurev.2019.03.005>
- [4] M. Wiswall and B. Zafar, "Determinants of college major choice: Identification using an information experiment," *The Review of Economic Studies*, vol. 82, no. 2, pp. 791-824, 2015. <https://doi.org/10.1093/restud/rdu044>
- [5] S. Anuvareepong, S. Phooim, N. Charoenprasoplar, and S. Vimongratana, "Course recommender system for student enrollment using augmented reality," in 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), 2017, pp. 212-217: IEEE. <https://doi.org/10.1109/iiiai-aa.2017.174>
- [6] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender systems handbook*: Springer, 2011, pp. 1-35. https://doi.org/10.1007/978-0-387-85820-3_1
- [7] D. Baneres and J. Conesa, "A life-long learning recommender system to promote employability," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 12, no. 06, pp. 77-93, 2017. <https://doi.org/10.3991/ijet.v12i06.7166>
- [8] M. El Mabrouk, S. Gaou, and M. K. Rtili, "Towards an intelligent hybrid recommendation system for e-learning platforms using data mining," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 12, no. 06, pp. 52-76, 2017. <https://doi.org/10.3991/ijet.v12i06.6610>
- [9] H. Lin, S. Xie, Z. Xiao, X. Deng, H. Yue, and K. Cai, "Adaptive recommender system for an intelligent classroom teaching model," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, no. 05, pp. 51-63, 2019. <https://doi.org/10.3991/ijet.v14i05.10251>
- [10] K. Pupara, W. Nuankaew, and P. Nuankaew, "An institution recommender system based on student context and educational institution in a mobile environment," in 2016 International Computer Science and Engineering Conference (ICSEC), 2016, pp. 1-6: IEEE. <https://doi.org/10.1109/icsec.2016.7859877>
- [11] S. Janpla and K. Kularbphetong, "The Recommended System For Students' Major Selection Based On Mining Educational Data," *TOJET*, 2016.
- [12] G. Abosamra and A. Faloudah, "Machine Learning Based Marks Prediction to Support Recommendation of Optimum Specialization and Study Track," *International Journal of Computer Applications*, vol. 975, p. 8887. <https://doi.org/10.5120/ijca2019918672>
- [13] J. Rajni and D. B. Malaya, "Predictive analytics in a higher education context," *IT Professional*, vol. 17, no. 4, pp. 24-33, 2015. <https://doi.org/10.1109/mitp.2015.68>
- [14] M. Y. S. Reddy and P. Govindarajulu, "College Recommender system using student preferences/voting: A system development with empirical study," *International Journal of Computer Science and Network Security*, vol. 18, no. 1, pp. 87-98, 2018.
- [15] D. Grewal and K. Kaur, "Developing an intelligent recommendation system for course selection by students for graduate courses," *Business and Economics Journal*, vol. 7, no. 2, 2016.
- [16] S. Shahab, "Next Level: A Course Recommender System Based On Career Interests," 2019. <https://doi.org/10.31979/etd.z4v2-k6gg>
- [17] N. B. Samrit and A. Thomas, "A Recommendation System for Prediction of Elective Subjects," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 5, no. 4, pp. 36-43, 2017. <https://doi.org/10.22214/ijraset.2017.4007>

- [18] A. Al-Badarenah and J. Alsakran, "An automated recommender system for course selection," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 3, pp. 166-175, 2016. <https://doi.org/10.14569/ijacsa.2016.070323>
- [19] B. Bakhshinategh, G. Spanakis, O. R. Zaiane, and S. ElAtia, "A Course Recommender System based on Graduating Attributes," in *CSEDU* (1), 2017, pp. 347-354. <https://doi.org/10.5220/0006318803470354>
- [20] K. Bhumichitr, S. Channarukul, N. Saejiem, R. Jiamthapthaksin, and K. Nongpong, "Recommender Systems for university elective course recommendation," in 2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2017, pp. 1-5: IEEE. <https://doi.org/10.1109/jcsse.2017.8025933>
- [21] E. J. Castellano and L. Martínez, "A web-decision support system based on collaborative filtering for academic orientation. Case Study of the Spanish Secondary School," *J. UCS*, vol. 15, no. 14, pp. 2786-2807, 2009.
- [22] Mission Vision And Goals. Available: <https://etec.gov.sa/ar/About/Centers/Pages/qiyas.aspx>
- [23] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in artificial intelligence*, vol. 2009, 2009. <https://doi.org/10.1155/2009/421425>
- [24] J. Bobadilla, F. Ortega, A. Hernando, and J. Bernal, "A collaborative filtering approach to mitigate the new user cold start problem," *Knowledge-based systems*, vol. 26, pp. 225-238, 2012. <https://doi.org/10.1016/j.knosys.2011.07.021>

10 Authors

Khlood Alshaikh is a Master student in information Systems Department in the Faculty of Computing and Information Technology at King Abdulaziz University, Jeddah, KSA, e-mail: khassanalshaikh@stu.kau.edu.sa.

Naela Bahurmuz is a Master student in information Systems Department in the Faculty of Computing and Information Technology at King Abdulaziz University, Jeddah, KSA, e-mail: nomarbahurmuz@stu.kau.edu.sa.

Ola Torabah is a Master student in information Systems Department in the Faculty of Computing and Information Technology at King Abdulaziz University, Jeddah, KSA, e-mail: otorabah@stu.kau.edu.sa.

Sarah Alzahrani is a Master student in information Systems Department in the Faculty of Computing and Information Technology at King Abdulaziz University, Jeddah, KSA, e-mail: salzahrani1331@stu.kau.edu.sa.

Zainab Alshingiti is a Master student in information Systems Department in the Faculty of Computing and Information Technology at King Abdulaziz University, Jeddah, KSA, e-mail: zainab.alshingiti@gmail.com.

Maram Meccawy is an Associate Professor at the Department of Information Systems, Faculty of Computer Science and Information Technology in King Abdulaziz University (KAU) in Saudi Arabia. Her research interests include: Adaptive Educational Hypermedia, User Modelling, LMS and MOOCs. e-mail: mmec-cawy@kau.edu.sa.

Article submitted 2020-08-17. Resubmitted 2020-09-26. Final acceptance 2020-09-29. Final version published as submitted by the authors.