CareerRec: A Machine Learning Approach to Career Path Choice for Information Technology Graduates

Hmood Al-Dossari
Information Systems Department
King Saud University
Saudi Arabia

Fawaz Abu Nughaymish
Information Systems Department
King Saud University
Saudi Arabia

Ziyad Al-Qahtani
Information Systems Department
King Saud University
Saudi Arabia

Mohammed Alkahlifah
Information Systems Department
King Saud University
Saudi Arabia

Asma Alqahtani
Information Systems Department
King Saud University
Saudi Arabia

Abstract-**Enterprises rely more and more on well-qualified and highly specialized IT professionals. Although the increasing availability of IT jobs is a good indicator for IT graduates, they nonetheless may find themselves confused about the most appropriate career for their future. In this paper, a recommendation system called CareerRec is proposed, which uses machine learning algorithms to help IT graduates select a career path based on their skills. CareerRec was trained and tested using a dataset of 2255 employees in the IT sector in Saudi Arabia. We conducted a performance comparison between five machine learning algorithms to assess their accuracy for predicting the best-suited career path among 3 classes. Our experiments demonstrate that the XGBoost algorithm outperforms other models and gives the highest accuracy (70.47%).**

Keywords-recommendation system; information technology graduates; machine learning; career selection

I. INTRODUCTION

Information Technology (IT) jobs continue to grow in several disciplines, such as cloud computing, cyber security, mobile applications and big data analytics [1]. Companies rely more and more on well-qualified and highly specialized IT professionals. Although the increasing availability of IT jobs is a good indicator for IT graduates, they nonetheless may find themselves confused about the most appropriate future career. Thus, there is a need to design and implement a system that can assist IT graduates in profession selection based on their skills. Recommendation systems have been a research topic for a long time [2]. They are simple algorithms that aim to provide the most relevant and accurate items to the user. A recommendation engine discovers data patterns in a data set by learning users' choices and produces outcomes that correlate to their needs and interests. The recommendee can make a huge difference in some industries by generating a huge amount of income, as with Amazon or Netflix. In this paper, we propose CareerRec, a recommendation system that relies on machine learning algorithms to help IT graduates choose their future career. The proposed system will recommend a career path based on some basic details about the IT graduate's skills, like soft skills (e.g. communication skills, logical thinking) and technical skills (e.g. programming languages, databases, UML, Spark, Networks). CareerRec can be used by senior students, job seekers, and employees. Senior students would benefit from the recommender's guidance about which course is most relevant to their preferences and interests. The recommender can also be utilized by job seekers to explore and identify the skills required by the IT market. Finally, employees can benchmark themselves against the market and recognize the skills they need to improve their positions.

To build CareerRec, data were collected from 2255 employees in the IT sector in Saudi Arabia, including their job titles and the skills they possess. The collected data have been used to train and test supervised machine learning algorithms to identify the most appropriate career for the IT graduate among 3 career paths (developer, analyst, and engineer). We compared the trained classifiers to assess their accuracy in predicting careers. Our experiments demonstrate that the XGBoost algorithm outperforms other models in the performance of career advice recommendation. The XGBoost algorithm gives the highest accuracy (70.47%) with the lowest error rate. All experiments were conducted using the Python programming language.

Corresponding author: Hmood Al-Dossari

II. BACKGROUND

IT plays an essential role in creating jobs in all industries. In the United States, for instance, more than 565,000 new IT jobs were created between 2001 and 2011, over 95 times more than the total employment in other occupations [3]. Furthermore, even over the recession, IT jobs grew much faster than non-IT jobs between 2007 and 2011, increasing by 6.8% and contributing almost $37 billion to the economy of the United States [4]. According to the US Bureau of Labor Statistics [5], software developers and programmers are expected to add 279,500 jobs by 2022, accounting for about 40% of new jobs in the computing and mathematics fields. Although the projected growth for information security analysts is smaller than for software developers and programmers, the rate of growth for information security analysts is expected to be 36.5% making this the fastest growing IT job. Demand for IT occupations stems from a number of factors, including the increased need for cybersecurity, the implementation of electronic medical records, and the increase in the use of mobile technology. According to [6], the 10 most in-demand IT jobs for 2020 are AI architect, business intelligence analyst, cloud architect, data professional, developer, DevOps engineer, helpdesk and desktop support professional, network/cloud administrator, network security professional, and system administrator. Saudi Arabia has experienced rapid economic growth that has been reflected in the increased adoption of Information and Communications Technologies (ICTs) in many aspects of business and government organizations [7]. The spending on ICT products and services has increased and reached 138.48 billion Saudi Riyals in 2017, which reinforces the ICT market in Saudi Arabia as the largest in the Middle East [7]. Many new jobs in IT are being created through growing industries that require advanced technical skills, putting pressure on local labor markets and technical education systems. In Saudi Arabia, the gap between supply and demand will nonetheless continue to expand [7]. Reducing the gap between the skills the IT market needs and the skills of IT graduates is essential in bridging the gap between supply and demand in the ICT sector.

III. RELATED WORK

Several methods use recommender systems to support decision making [2]. Collaborative filtering methods are widely used in recommender systems [8, 9]. Generally, there are two collaborative filtering approaches: user-based [10] and item-based [11]. The user-based collaborative filtering approach defines the similarity between two users based on the services or products they commonly used or bought. The item-based collaborative filtering approach, on the other hand, defines the similarity between services or products instead of users. Authors in [12] propose a career recommender system to support the decisions of senior high school students about the career they will pursue. To do so, they collected data from 716 senior high school students in the Philippines. The proposed recommender was built upon a fuzz-based engine to provide suitable recommendation. They presented 72 rules for the fuzzy model and their system produces reasonable results for making decisions. A recommender system was proposed in [13], in order to help perspective students select suitable IT companies

in Nigeria. The data were collected through an online questionnaire, with 200 respondents. This employed a collaborative filtering recommendation approach, using the C4.5 algorithm to classify the data and generate a decision tree model from the training data set (with 78.84% accuracy). The developed model was used as a knowledge base for a very beneficial front-end web application where students can enter their preferences and view company recommendations.

Authors in [14] developed a model to provide recommendations for job seekers by matching their profiles with persons with similar profiles (e.g. educational background, professional skills). The data were collected through a Google survey distributed on social media in Pakistan. This study used the Apriori algorithm to mine and extract association rules from the collected data. The algorithm was implemented in R Studio and 62 association rules were generated to support the recommendations. Recommender systems for educational guidance were developed in [15] to support course selection by students. The proposed feasible predictions for student course selection were based on their marks and choice of job interests. The targeted population for this study consisted of students eager to join fields like engineering, medical, commerce, arts, etc. The authors collected data from 1500 students in India. They used clustering techniques such as the K-Means Clustering algorithm to find structures and relationships within the data. Then, they used an association rule to inspect the associations linking the subgroups. This process was then applied to determine the student characteristics that align with individual characteristics. Finally, classification based on fuzzy set theory and rough sets was applied. This system suggested suitable information based on courses, jobs, and activities to support a student's decision. Finally, students were able to make a final decision related to their studies. They filled out a feedback form students and satisfaction was expressed in 95% of the cases. Authors in [13] proposed the use of data mining techniques to predict students' final GPA based on their grades in previous courses. The data were collected from 236 transcripts of female students who graduated from Computer Sciences College at King Saud University, Saudi Arabia in 2012. A decision tree was implemented using WEKA tools and then was used to predict students' GPA based on their grades in mandatory courses. This identified the most important courses in the study plan, which had the larger impact on the students' final GPAs.

Authors in [17] investigated the impact of assessment grades and online activity data in the Learning Management System (LMS) on students' academic performance. The dataset consisted of 241 records of undergraduate students from six different courses delivered from 2017 to 2019. The data were obtained from the Deanship of E-Learning and Distance Education at King Abdulaziz University. Students' data included assessment grades and blackboard activity data. The classification algorithms were implemented using WEKA tools. Five algorithms were used in this study: decision tree, random forest, sequential minimal optimization, multilayer perceptron, and logistic regression. The results revealed that the random forest algorithm performs better for predicting student academic performance, followed by the decision tree. The purpose of the authors in [18] was to investigate the acceptance

level of the LMS among students in Payamnoor and Farhangian universities. The study designed a questionnaire that was answered by 200 students, and the results show that most students, regardless of gender, age, and department, were satisfied with the usage of the LMS. However, student's grades seem to play a significant role in the level of satisfaction from the LMS. Authors in [19] proposed a model for the development of employees' learning (career path) in industrial enterprises. They used a descriptive survey method and field research, the statistical population consisting of all 110 the employees of Dana Baspar's enterprises. The result shows that the influence of workshop and experimental skills gained by apprenticeship, the effect of training by holding meetings and seminars with experts, and the experience acquired during work are all significant and positive on organizational productivity. The mediation variable of professional skills exists, but the influence of classic and academic training is not positive and significant on organizational productivity, considering the existence of mediation variables. Authors in [29] proposed a weighting method that can be used to combine two or more social context factors in a recommendation engine that leverages an Exponential Random Graph Model (ERGM) based on historical network data. In this paper and in the same (educational) context, we propose a recommender to support an IT graduate in selecting his/her future career. The main idea is to calculate skill similarities between the candidate and current employees.

IV. RESEARCH METHODOLOGY

To develop the proposed recommender, the methodology suggested in this research includes the following phases: data collection, data pre-processing, and classification. The flow diagram of our proposed methodology is illustrated in Figure 1, and the implementation and experimental study will be described below.
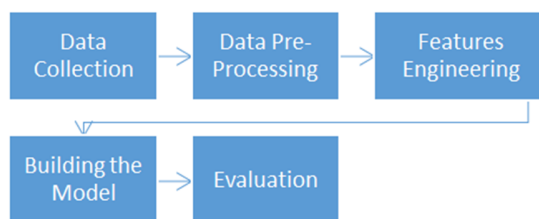

Fig. 1.   Flow diagram of the research methodology.

A. Data Collection

In this phase, a survey to collect data from current IT employees in the Saudi market was designed. The survey consisted of several parts such as demographic data, academic professions, and soft and technical skills. The survey was a combination of open and closed questions (see Appendix). The survey was then published online, targeting IT employees in Saudi Arabia. Focus was given on collecting data from members who wrote on their LinkedIn profile that they work in IT jobs. Several enterprises, whether public or private, were covered in the survey, either by direct communication or via email (e.g. universities, Elm company, Saudi Telecom, Mobily, SABIC, Aramco, CITC, KACST, and other companies). The

data collection stage ended with 2255 responses, and after data filtering, the dataset used in the experiments contained 2167 records.

B. Data Pre-processing

In this phase, pre-processing was applied to the collected data to make them suitable for the model-building phase. First, duplicated rows were removed and missing values were assigned. The authors communicated directly with employees via their emails for data imputations or data corrections. Then, some inconsistencies with data have been fixed, for example, the number of years has been written as 2 instead of "two". Also, the inconsistency between university names and departments has been addressed (e.g. Information Systems instead of IS, King Saud University rather than KSU). Finally, we conducted several data transformations as follows:

- In the designed survey, the participants were asked to fill in 20 skills (10 soft skills and 10 technical and programming skills). The participants were asked to evaluate their level in a given skill as none, low, moderate, or high. Programming skills, on the other hand, were designed as a text-free field where the participants entered the list of programming languages they know. After a quick exploration, we noticed that the programming skill included 49 unique programming languages, so we decided to create 49 additional attributes, where each represented a given language (e.g. Python) and is "1" if the participant knows the language and "0" if not.

- The academic profession has a feature identifying the main major for the respondents (Your Major/Specialization). It had almost 251 unique values and we followed the ACM.org curricula recommendations [21] to transform the specialization into its root. According to [21], there are 5 categories for computing specialists: Computer Science, Information Systems, Information Technology, Computer Engineering, and Software Engineering. It has been observed that some majors are actually not relevant to computing, although the respondents are working in the Saudi IT market. Thus, in addition to these 5 categorizations, we added one more category for non-computing majors. At the end of this exercise, the values of the "Your Major/Specialization" feature were transformed into one of the six identified categories, shown in Figure 2.
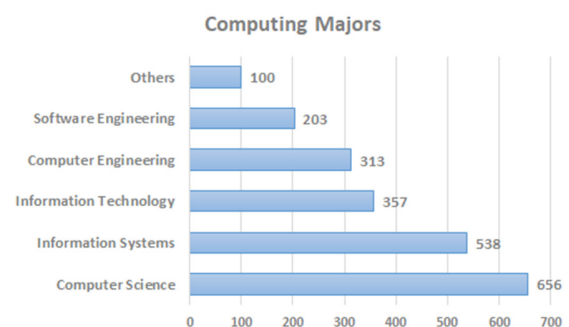

Fig. 2.   Distribution of computing majors.

- Finally, the "job title" feature had 565 unique values. Cleaning up this feature was not trivial because the participants wrote job titles using different names for the same job. Also, sometimes there were spelling mistakes in the job titles. By following [22, 23], we transformed the values of "job title" into 37 distinct job titles (Figure 3). For the purposes of this study, we classified these 37 jobs into 4 groups (or career paths) shown in Table I. The last group (Others) represents job titles not directly relevant to computing (i.e. managerial or support jobs).
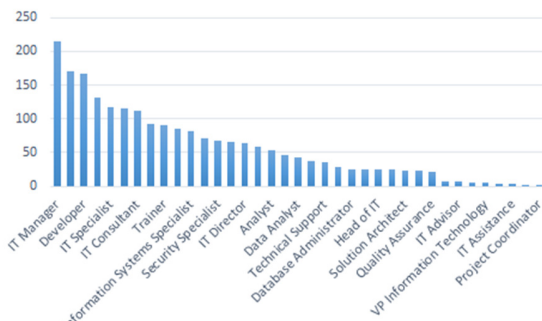
Fig. 3.   Distribution of job titles.

TABLE I.   IT JOBS CLASSIFICATION

Career path	Job titles
Analyst (10 jobs)	Analyst, Systems Analyst, Business Analyst, Data Analyst, Application Analyst, Solution Architect, Database Administrator, IT Consultant, IT Director
Developer (8 jobs)	Developer, Programmer, Business Developer, Information Systems Specialist, Security Specialist, Quality Assurance Specialist, IT Specialist, Software Engineer
Engineer (4 jobs)	Engineer, System Engineer, Network Engineer, System Administrator, Technical Support Specialist
Others (14 jobs)	Chief Executive Officer, Head of IT, IT Advisor, IT Assistant, IT Officer, IT Supervisor, Project Coordinator, Researcher, Team Leader, UX/UI Designer, VP of Information Technology, Trainer, IT Manager, Computer Operator

C. Feature Engineering

In this phase, some irrelevant features that did not contribute to predicting the best career path for the IT graduate were removed. Features such as nationality, gender, email, job location, work sector, and work domain were removed from the data. We used label encoding with categorical variables (e.g. soft skill, technical skill, degree, and major). Encoding is one way of using categorical predictor variables in various kinds of estimation models. The target categorical variable is encoded with a value between 0 and n, and the code can be easily decoded by using the same value.

D. Building The Model

The aim of this study was to develop a recommendation system to help IT graduates select the career best suited for their skills. The research problem has been defined as a multi-classification problem where the potential graduate enters his own skills and the system suggests him/her one of three career paths (Developer, Analyst, Engineer). To do this, the scikit-learn library in Python was used to implement several machine learning algorithms. Specifically, five machine learning algorithms were used to build the model:

- K-Nearest Neighbors (KNN): It is a simple and easy-to-implement supervised machine learning algorithm that can be used to solve classification and regression problems. The KNN algorithm stores all available cases and classifies new cases based on a similarity measure. The KNN classifies a new case by a majority vote of its neighbors [24].

- Decision Tree (DT): It is a supervised machine learning technique. In DT, each internal node represents a "test," each leaf node represents a class label, and the path from the root to leaves represents the classification rules [25].

- Bagging meta-estimator: This is an ensemble meta-algorithm combining predictions from multiple decision trees through a majority voting mechanism [26].

- Gradient Boosting: It is a machine learning technique for regression and classification problems that produces a prediction model in the form of an ensemble of weaker prediction models, typically DTs. Gradient boosting employs a gradient descent algorithm to minimize errors in sequential models [27].

- XGBoost: This is a DT-based ensemble Machine Learning algorithm that uses a gradient boosting framework [28]. XGBoost is an optimized gradient boosting algorithm that uses parallel processing, tree-pruning, missing value handling, and regularization to avoid overfitting or bias.

V.   EXPERIMENTS AND RESULTS

In this section, we evaluate and compare the five classifiers.

A. Dataset Description

The data used in this study were collected by a survey distributed among IT employees in the Saudi market. The participants in the survey included employees, leaders, and decision-makers in the field of IT in many government and private sectors. The data were collected during the period from February 2018 to October 2018 and the number of respondents was 2225. The main characteristics of the respondents are shown in Figure 4.
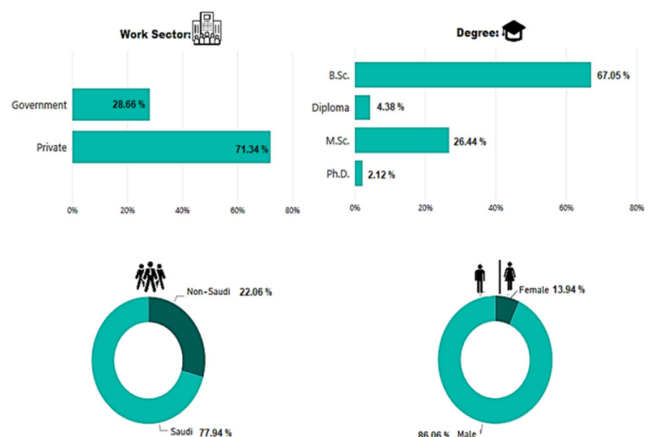

Fig. 4.   Characteristics of the respondents.

B. Descriptive Analytics

In this phase, descriptive analytics were conducted to explore and understand the IT market in Saudi Arabia. More specifically, we wanted to identify the most important skills that employees have in this sector. At first, we explored the skills of IT employees who have spent at most five years (entry level) in the labor market. The knowledge of these skills is important for senior students and job seekers. Figures 5 and 6 illustrate the soft and the technical skills of the entry-level employees in the Saudi market. We suppose an employee posseses a skill if he identified himself as "high" in the targeted skill.
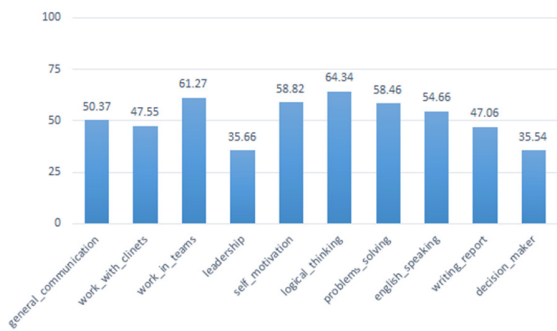
Fig. 5. Percentage of employees identifying themselves as "high" in each soft skill.
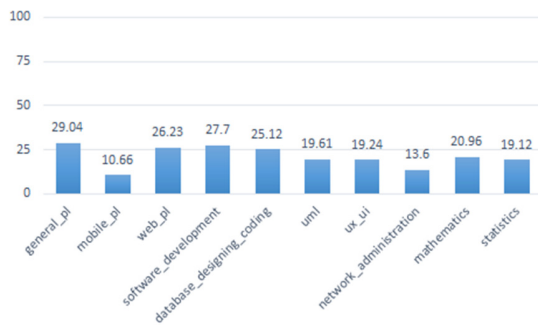
Fig. 6. Percentage of employees identifying themselves as "high" in each technical skill.

From Figures 5 and 6, in general, it is noticeable that the percentage of employees possessing soft skills is much higher that the percentage that possess technical skills. It is obvious that the percentage with the highest ranking technical skill is less than the percentage with the lowest ranking soft skill (29.04% in general programming languages is the highest score in soft skills, and 35.54% in decision making skill is the lowest score in technical skills). One possible explanation is that it is easier for IT employees to evaluate whether they have a given technical skill. For example, the percentage of employees who identified themselves as "high" in mobile programming language skills (10.66%) is close to the percentage of those with programming skill in the language Swift (13.24%). In contrast, personal skills are difficult to measure, so we find that employees may exaggerate these skills in a way that does not reflect reality. In Figure 5, it is obvious that the most important

soft skills for entering the Saudi IT market are logical thinking, working in teams, self-motivation, and problem solving. The most important technical skills, are general programming, software development, web programming, and database designing and coding. These 8 skills can be considered the minimum required skills to enter the Saudi IT market. Figure 7 illustrates the top 7 programming languages that employees in the Saudi markets know. Java is at the top of this list at almost 67%, which is expected since computing departments at Saudi universities teach this language in their programming courses. Similarly, some departments teach C++ and PHP in their academic curriculum.
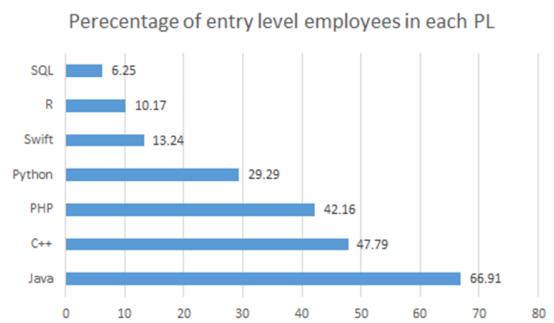

Fig. 7. Percentage of entry-level employees with each PL.

C. Predictive Analytics

Five machine learning algorithms were trained and tested to predict the best suited career path for the computing graduates. KNN, DT, Bagging meta-estimator, Gradient Boosting, and XGBoost were trained using a dataset of 1707 employees to predict the most suitable career among three classes: Analyst, Developer, and Engineer. The "Others" class was omitted because the jobs in this class require long experience and specific requirements that do not exist among fresh graduates. These algorithms were trained using 27 features: 10 soft skills (Figure 5), 10 technical skills (Figure 6), and 7 programming language skills (Figure 7). During data preparation, we observed that respondents mentioned 49 programming languages. To build the model, we chose the programming languages mentioned by more than 5% of the respondents. The result is shown in Figure 8.
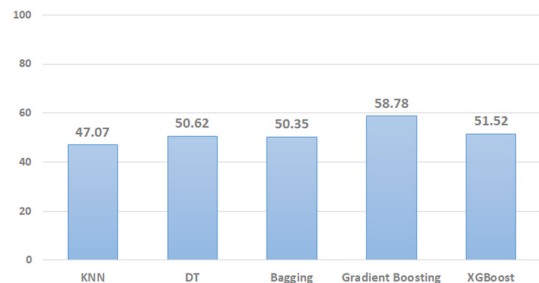

Fig. 8. Accuracy of the 5 ML algorithms in the prediction of the best career path.

The result of this phase shows that the Gradient Boosting algorithm achieved the highest accuracy score at 58.78%,

followed by XGBoost at 51.52%. From the confusion matrix in Figure 9(a), it can be observed that the Engineer class had the lowest accuracy (58%), as a large number of participants were incorrectly classified to this class (19% were classified as Developers and 23% were classified as Analysts). This occurred because this class is the least represented class in this experiment. In the other words, the dataset in this experiment was imbalanced (Analyst 38.49%, Developer 42.59%, Engineer 18.92%).
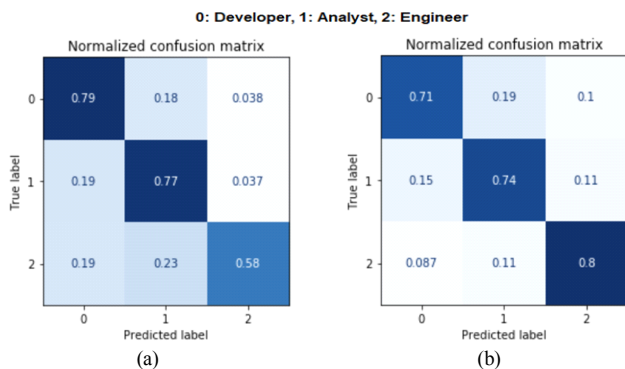

Fig. 9.   Confusion matrix of XGBoost algorithm: (a) Imbalanced dataset, (b) balanced dataset.

To address imbalanced datasets, an oversampling technique was used to replicate the minority class (Engineer class in our case) in order to increase its cardinality. The new distribution of our classes after replicating the Engineer class was almost balanced (Analyst 32.36%, Developer 35.81%, Engineer 31.81%). The five ML algorithms were trained and tested again on the new dataset, and the result is shown in Figure 10.
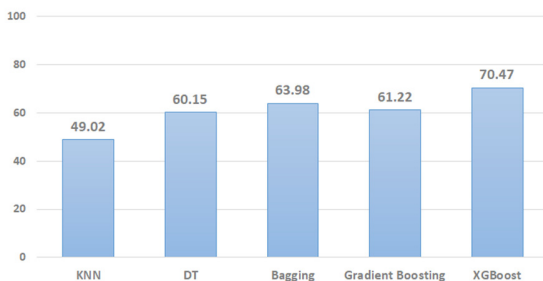

Fig. 10.   Accuracy of the 5 ML algorithms with a balanced dataset.

It is obvious that XGBoost achieves the highest accuracy at 70.47%, followed by the Bagging algorithm at 63.98%. From Figure 9(b), it can be noticed that the classification accuracy for the Engineer class has improved significantly (from 58% to 80%), and a slight negative impact can also be observed on the accuracy of the classifications of the Developer and Analyst classes.

VI.   DISCUSSION

The scope of this study was the prediction of the best career path for computing graduates as an application of ML algorithms in the field of human resources. This application is not only useful for the computing graduate but can also be used by employers to determine the applicant's suitability for a specific job position. The low accuracy achieved in our experiments can be justified by the small dataset that was used. Collecting more data about IT employees in the Saudi market may help boost classification accuracy. One reason for the low classification accuracy is the method of data collection, which was direct asking the employees about their level of skill acquisition. This method is affected by the credibility and accuracy of the employee's self-evaluation. It might be useful to investigate and use other approaches to define and recognize employees' skills rather than collecting them directly from the employees. One potential approach to improve the performance and reliability of the proposed model is to acquire employees' skills indirectly from their colleagues and supervisors. Moreover, additional data can be collected from the employees themselves to make sure they fit with their current jobs (i.e. that they are capable at and satisfied with the current position). On the other hand, it might be better to obtain the technical skills of IT graduates from their academic records rather than asking them directly. In general, such models offer significant challenges in collecting data and in credibility. It must be emphasized that the accuracy and reliability of the collected data will directly affect the validity of any proposed model.

VII.   CONCLUSION

In this paper, we have proposed CareerRec, a recommendation system that helps IT graduates select the career that best suits their skills. To do so, five machine learning algorithms were trained and tested using a dataset of employees in the Saudi IT sector. The experimental results showed that the XGBoost algorithm outperforms the other models in supporting career path selection for IT graduates.

The performance of the proposed system can be improved by collecting more data. More algorithms, such as deep learning models, can be considered to improve the accuracy of the proposed system. Also, an alternative approach to measure IT employees' skills might be considered, to provide more accurate and trustworthy training data.

APPENDIX

The questionnaire:

Diploma or Bachelor university: *

Your answer

Your Major/Specialization: *

Your answer

*

	Government	Private
Work Sector:	○	○

In which sector you are working in: *

○ Education
○ Telecom
○ Healthcare
○ Consulting
○ IT
○ Banking
○ Other:

Years of Experience: *

Your answer

Job title: *

Your answer

Job Location: *

Choose ▾

*

	Less than 10.000	10.000 - 20.000	20.000 - 30.000	More than 30.000
Salary range:	○	○	○	○

BACK   NEXT   Page 2 of 4

Personal skills: *

	None	Low	Medium	High
Self-motivation for Continued Lifelong Learning	○	○	○	○
Dealing with Co-workers/Clients	○	○	○	○
General Communication Skills	○	○	○	○
Judging and making decision	○	○	○	○
Writing English Reports	○	○	○	○
Problems solving	○	○	○	○
Working in Teams	○	○	○	○
Logical thinking	○	○	○	○
Speaking English	○	○	○	○
Leadership	○	○	○	○

BACK   NEXT   Page 3 of 4

Functional skills: *

	None	Low	Medium	High
General programming language	○	○	○	○
Web programming langage	○	○	○	○
Mobile programming langage	○	○	○	○
Database: Coding - Designing	○	○	○	○
Software development	○	○	○	○
Linux Application Development	○	○	○	○
Unified Modeling Language (UML)	○	○	○	○
User Interface UI - UX	○	○	○	○
Network Administration	○	○	○	○
Information Security	○	○	○	○
Cloud Computing	○	○	○	○
Big Data Analytics (Hadoop/Spark)	○	○	○	○
Machine learning technique	○	○	○	○
Mathematics	○	○	○	○
statistics - probability	○	○	○	○
Technical support	○	○	○	○

Knowledge of programming languages: *

☐ java
☐ c++
☐ python
☐ PHP
☐ R
☐ swift
☐ Other:

Your email if you are interested to get the results:

Your answer

REFERENCES

[1] C. M. Veneri, *Here Today, Jobs of Tomorrow: Opportunities in Information Technology.* Bureau of Labor Statistics, 1998.

[2] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, pp. 109–132, Jul. 2013, https://doi.org/10.1016/j.knosys.2013.03.012.

[3] L. Csorny, "Careers in the Growing Field of Information Technology Services," *Beyond the Numbers*, vol. 2, no. 9, Apr. 2013.

[4] L. Stewart, "Looking for Jobs? Look to IT in 2012 and Beyond," *Information Technology and Innovation Foundation*, 2012.

[5] D. Richards and D. Terkanian, "Occupational employment projections to 2022," *Monthly Labor Review, U.S. Bureau of Labor Statistics*, Dec. 2013, https://doi.org/10.21916/mlr.2013.41.

[6] S. K. White, "The 10 most in-demand tech jobs for 2020 — and how to hire for them," *CIO*, Jan. 10, 2020. https://www.cio.com/article/3235944/hiring-the-most-in-demand-tech-jobs-for-2018.html (accessed Dec. 12, 2020).

[7] *ICT Report: ICT Workforce in the Kingdom of Saudi Arabia.* Riyadh, Saudi Arabia: CITC Publications, 2015.

[8] H. Ma, I. King, and M. R. Lyu, "Effective missing data prediction for collaborative filtering," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, Jul. 2007, pp. 39–46, https://doi.org/10.1145/1277741.1277751.

[9]   P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," in *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, New York, NY, USA, Oct. 1994, pp. 175–186, https://doi.org/10.1145/192844.192905.

[10]   J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, San Francisco, CA, USA, Jul. 1998, pp. 43–52.

[11]   B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*, New York, NY, USA, Apr. 2001, pp. 285–295, https://doi.org/10.1145/371920.372071.

[12]   M. C. B. Natividad, B. D. Gerardo, and R. P. Medina, "A fuzzy-based career recommender system for senior high school students in K to 12 education," *IOP Conference Series: Materials Science and Engineering*, vol. 482, Mar. 2019, Art. no. 012025, https://doi.org/10.1088/1757-899X/482/1/012025.

[13]   A. O. Ogunde and J. O. Idialu, "A recommender system for selecting potential industrial training organizations," *Engineering Reports*, vol. 1, no. 3, 2019, Art. no. e12046, https://doi.org/10.1002/eng2.12046.

[14]   M. Kiran, H. Asim, and M. T. Hassan, "Career and Skills Recommendations using Data Mining Technique: Matching Right People for Right Profession, in Pakistani Context," *VFAST Transactions on Software Engineering*, vol. 7, no. 1, pp. 33-41–41, Dec. 2018, https://doi.org/10.21015/vtse.v13i3.510.

[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, 2015, Art. no. 1000209, https://doi.org/10.4172/2151-6219.1000209.

[16]   M. A. Al-Barrak and M. Al-Razgan, "Predicting Students Final GPA Using Decision Trees: A Case Study," *International Journal of Information and Education Technology*, vol. 6, no. 7, pp. 528–533, 2016, https://doi.org/10.7763/IJIET.2016.V6.745.

[17]   A. Alhassan, B. Zafar, and A. Mueen, "Predict Students' Academic Performance based on their Assessment Grades and Online Activity Data," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 4, Jun. 2020, https://doi.org/10.14569/IJACSA.2020.0110425.

[18]   P. Shayan and E. Iscioglu, "An Assessment of Students' Satisfaction Level from Learning Management Systems: Case Study of Payamnoor and Farhangian Universities," *Engineering, Technology & Applied Science Research*, vol. 7, no. 4, pp. 1874–1878, Aug. 2017, https://doi.org/10.48084/etasr.1041.

[19]   G. C. Keshvari, J. M. Faraz, N. Safaie, and S. J. H. Nedjad, "A Model For the Development of Employees' Learning (Career Path) in Industrial Enterprises," *Engineering, Technology & Applied Science Research*, vol. 8, no. 1, pp. 2427–2432, Feb. 2018, https://doi.org/10.48084/etasr.1687.

[20]   H. Al-Ballaa, H. Al-Dossari, and A. Chikh, "Using an Exponential Random Graph Model to Recommend Academic Collaborators," *Information*, vol. 10, no. 6, p. 220, Jun. 2019, https://doi.org/10.3390/info10060220.

[21]   M. Sahami, S. Roach, E. Cuadros-Vargas, and R. LeBlanc, "ACM/IEEE-CS computer science curriculum 2013: reviewing the ironman report," in *Proceeding of the 44th ACM technical symposium on Computer science education*, New York, NY, USA, Mar. 2013, pp. 13–14, https://doi.org/10.1145/2445196.2445206.

[22]   *International Standard Classification of Occupations 2008 (ISCO-08): Structure, group definitions and correspondence tables*. Geneva, Switzerland: ILO, 2012.

[23]   *The Unified Saudi Occupational Classification*. Riyadh, Saudi Arabia: General Authority for Statistics, 2019.

[24]   D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, no. 1, pp. 37–66, Jan. 1991, https://doi.org/10.1007/BF00153759.

[25]   J. R. Quinlan, *C4.5: Programs for Machine Learning*, 1st ed. San Mateo, CA, USA: Morgan Kaufmann, 1992.

[26]   L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, Aug. 1996, https://doi.org/10.1007/BF00058655.

[27]   J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, Feb. 2002, https://doi.org/10.1016/S0167-9473(01)00065-2.

[28]   T. Chen and T. He, *xgboost: eXtreme Gradient Boosting*. 2020.