CareProfSys - Combining Machine Learning and Virtual Reality to Build an Attractive Job Recommender System for Youth: Technical Details and Experimental Data

Maria-Iuliana Dascalu[1], Andrei-Sergiu Bumbacea[1], Ioan-Alexandru Bratosin[1], Iulia-Cristina Stanica [1] and Constanta-Nicoleta Bodea[2,3]

[1] National University of Science and Technology POLITEHNICA Bucharest, Splaiul Independenței 313, Bucharest 060042, Romania
[2] Department of Economic Informatics and Cybernetics, Bucharest University of Economic Studies, Calea Dorobantilor 13-15, Bucharest 010552, Romania
[3]“COSTIN C. KIRITESCU”, National Institute for Economic Research, Romanian Academy, Calea 13 Septembrie 13,Bucharest 050711, Romania
maria.dascalu@upb.ro, andrei.sergiu1999@gmail.com, ioan.bratosin@gmail.com, iulia.stanica@upb.ro, bodea@ase.ro

Abstract. The current article presents CareProfSys - an innovative job recommender system (RS) for youth, which integrates several emergent technologies, such as machine learning (ML) and virtual reality on web (WebVR). The recommended jobs are the ones provided by the well-known European Skills, Competences, Qualifications, and Occupations (ESCO) framework. The machine-learning based recommendation mechanism uses a K-Nearest Neighbors (KNN) algorithm: the data needed to train the machine learning model was based on the Skills Occupation Matrix Table offered by ESCO, as well as on data collected by our project team. This two-source method made sure that the dataset was strong and varied, which made it easier for the model to make accurate recommendations. Each job was described in terms of the features needed by individuals to be good professionals, e.g., skill levels for working with computers, constructing, management, working with machinery and specialized equipment, for assisting and caring, for communication collaboration and creativity are just a few of the directions considered to define a profession profile. The recommended jobs are described in a modern manner, by allowing the users to explore various WebVR scenarios with specific professional activities. The article provides the technical details of the system, the difficulties of building a stack of such diverse technologies (ML, WebVR, semantic technologies), as well as validation data from experiments with real users: a group of high school students from not so developed cities from Romania, interacting first time with modern technologies.

Keywords: Recommender System, Machine Learning, Virtual Reality on Web

1 Introduction

Career coaching is essential for preparing young individuals with the information and skills required to navigate a job market that is complicated and constantly evolving. Also, insufficient career advising resources in developing countries, such as Romania, have made the transition from education to job difficult for them. There are certain challenges encountered by young adults in developing countries that require our attention and discussion. Young adults in Romania, particularly those living in rural regions, have limited access to career counseling facilities. This lack of accessibility hinders their chances for long-term success in the job market by preventing them from making educated decisions on their education and future pathways [1]. At the same time, the job search process may be time-consuming, inefficient, and irritating for both job seekers and employers. To solve these difficulties, researchers have started creating RSs that employ deep reinforcement learning for online advertising. These systems strive to enhance the entire job search process by giving tailored job recommendations based on an individual's likes and talents.

This article presents a RS(RS) to enhance the efficacy and efficiency of job choosing for young individuals: the recommendations are made using a machine-learning (ML) technique and the provided results do not contain only a text-based description of a job, but a gamified virtual reality (VR) scenario of activities specific to that job, thus allowing the young users to better understand whether that job is suitable or not for them. To implement ML, accurate data about the features of practitioners of a certain job is needed. We used the Skills Occupation Matrix Table offered by European Skills, Competences, Qualifications, and Occupations (ESCO) framework [2], but also data provided by Romanian practitioners, thus we claim that the recommendations are useful in developing countries, such as Romania, but also aligned with the European landscape. Experimental data was collected from 27 technological high school students from Romania, participants at a summer school at National University of Science and Technology POLITEHNICA Bucharest. The young students, aged between 16 and 18 years old, decided to follow engineering, but were still undecided what engineering specialization to choose. Thus, they found the recommendations made by our system to be enlightening. Our article presents the ML-based recommendation technique, in the context of other RSs and underlines the efficiency of embedding WebVR in a career guidance system.

2 Job Recommender Systems

2.1 Recommender Systems: Usage and Typologies

A RS offers consumers customized on-line service or product tips to combat the expanding trouble of online information overload as well as to boost consumer connection monitoring. There are three main types of RSs: content-based, collaborative filtering, and hybrid ones. The purpose of content-based filtering systems is to recommend items based on the collected knowledge of users. This approach concentrates on matching individual interests with item attributes, so it is essential that the system includes one of the most essential item features. Concern number one prior to creating a system must be to determine each user's recommended features. This can be completed through a combination of the two strategies. Users are initially provided with a checklist of functions from which they can select those that most ignite their passion. Second, the recommendation algorithms keep a document of all items formerly chosen by the user, which acts as the premise for the customer's behavioral data. The customer's account is based on their preferences, inclinations, as well as options, which inevitably affect their rankings [3].

Collaborative filtering recommenders (CF RSs) require a set of things based on the individual's previous choices. Each object and user are described by an embedding or attribute vector, which places both the objects and the individuals in a similar embedding location. When suggesting a certain product to the primary user, the viewpoints of other customers are taken into consideration. It keeps an eye on the activities of all users to identify which item is the most popular. When suggesting an item to the main customer, it likewise relates comparable individuals based on their shared preferences and behavior toward a comparable product [3]. There are two types of collaborative filtering approaches. Memory-based Collaborative Filtering is an approach that determines the similarity between users or items by utilizing past user data, specifically rankings. The primary goal of this technique is to quantify the similarity between users or items, identify similar ratings, and recommend hidden items accordingly. Model-based approaches use ML models to predict and rank interactions between users and unengaged items. Using algorithms such as matrix factorization, deep learning, and clustering, etc., these models are trained utilizing the interaction information already available from the interaction matrix [4].

When making recommendations, a hybrid method mixes techniques from both collaborative and content-based filtering.

Understanding the differences between the types of RSs helps developers choose the best approach, as seen in Table 1.

Table 1. Criteria of choosing the RS type.

Criteria of choosing	Collaborative RSs	CB RSs	Hybrid RSs
Data used	User-item interaction data (e.g., clicks, purchases, ratings)	User-item interaction data (e.g., clicks, purchases, ratings)	Both user-item interaction data and item content data
User & Item cold start	Sensitive (new users & items have no interaction data)	Less sensitive (can use user preferences and item features to recommend new data)	Can handle cold start problems by leveraging content-based filtering component
Scalability	Can be computationally expensive for large user-item matrices	Can be computationally expensive for large user-item matrices.	Depends on the specific combination of methods used

Diversity of provided recommendations	Can be high	Can be lower	Can be adjusted
Personalization	High	Moderate	Can be high if the combination of methods is optimized

CF RSs are used in music streaming, movies recommendation and e-commerce. CB RSs are also used in movies recommendations, but more in news articles' platforms and job portals, while hybrid RSs have a wide variety of exploitation cases.

2.2 Recommender Systems used in Career Path Profiling and Jobs' Searching

In recent years, one may have witnessed a boom in the popularity of computer-based applications that assist individuals in the process of job searching. CareerExplorer by PathSource [5] uses ML algorithms to match users with careers based on their skills, interests, values, and work preferences. Users take an online assessment, and the platform generates personalized career suggestions along with relevant educational and job opportunities. The major limitation of the system is the fact that it depends on self-reported data, and the assessment process can be time-consuming. O*NET Interest Profiler [6] is a free tool that helps users identify their interests and match them to potential careers. Users answer questions related to their likes and dislikes, and the profiler suggests careers based on the user's interests. The system is limited to U.S. labor market data and may not be as engaging as other platforms. Pymetrics [7] uses neuroscience-based assessments, such as games and puzzles, to measure users' cognitive and emotional traits. The main drawbacks are: it is based on a relatively new and untested method, has a limited career database, and may not be suitable for all users due to the game-based assessment format. Jobscan [8] is primarily focused on optimizing resumes and LinkedIn profiles for specific job postings. However, it also offers a career path exploration feature by suggesting related job titles and industries based on the user's current resume or LinkedIn profile. Jobscan is limited in terms of career exploration, relies on existing resume or LinkedIn profile data, and may not cater to users seeking a significant career change. MyNextMove [9] allows users to search for careers based on keywords, browse careers by industry, or take an interest-based assessment (similar to O*NET Interest Profiler) to receive career suggestions. It is limited to U.S. labor market. Although all the systems have strengths and limitations, none is suitable for young people from European Union, especially from developing countries such as Romania. That is why CareProfSys, our recommended system, has a place in this plethora of career path guidance and job recommendation systems, by combining a ML hybrid recommender system with other emerging technology, very appealing to young people, VR on Web. VR technologies are increasingly used for developing professional capabilities, as required by jobs in different industries, e.g., in civil engineering [10], in medicine [11] or science [12].

3 CareProfSys: a Job Recommender System for Youth

Following registration and login, a user must answer a quiz about one' s skills and then job recommendations are listed. After that, the user can access a WebVR scenario of basic activities typical for that job. By exploring this game-like scenario, the young user may better understand the details of a certain recommendation. Premium users have the possibility to browse the entire list of skills needed for a certain job, thus they can improve one's CV. In Fig. 1, snippets from CareProfSys system are provided: a list of recommended jobs and a scenario in VR for the profession of network specialist.

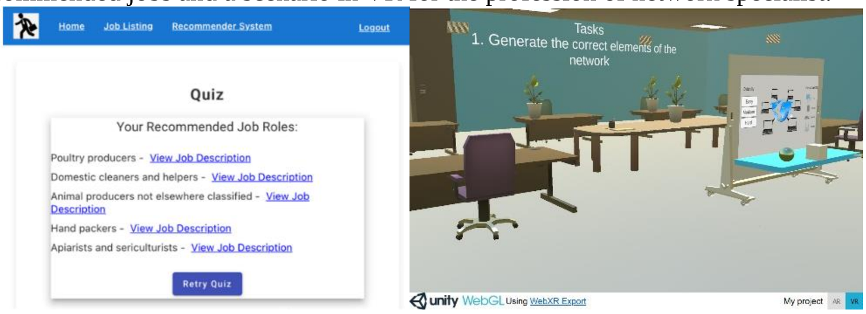

Fig. 1. Snippets from CareProfSys Recommender System.

3.1 Technical Details

Various technologies were used to make the recommender system work. The Angular framework was used to make the user interface for the recommender system. Angular is a strong tool for making dynamic, single-page web apps, and it helped make the user experience easy to use and responsive [13]. Flask is a simple, lightweight web platform for Python that was used to make a web server that the Angular application could talk to [14]. Python was used to build the core of the recommender system. The Python code was written so that it could read CSV files, preprocess the data, train the ML model, and then use this model to make job suggestions based on what the user said [15]. Specific packages of Unity Engine allowed us to execute a VR application directly from a web-browser: an application can be hosted on a web-browser if the build type of the application is WebGL [16]. WebGL is a Javascript API meant for rendering 3D graphics without the aid of additional plugins. WebXR Exporter is a Unity package that allows the development and build of VR applications in WebGL format compatible with multiple browsers such as Mozilla Firefox, Google Chrome, Microsoft Edge on Windows, Oculus Browser and Firefox Reality on Oculus Quest. The basic functionalities such as movement, rotation and interaction were handled using the VRTK Tilia packages which contain a collection of common functionalities meant for VR environments. Due to the WebGL format, the application is also compatible with multiple models of VR equipment as successful tests were done with HTC Vive Cosmos Elite, Oculus Rift and Oculus Quest. MongoDB, a JSON database, was also used: the Angular application sends user data to the Flask server in JSON format, which is further saved in the database and the Flask server responded with job recommendations in the same format [17].

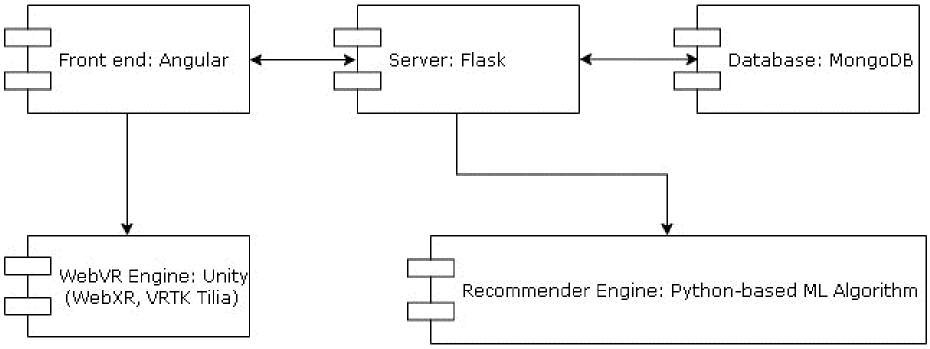

Fig. 2. Architecture of CareProfSys Recommender System.

3.2 Machine-learning based Recommendation Algorithm

To execute the ML algorithm, several steps had been taken, such as: data collection, data preprocessing and cleaning, feature extraction and engineering, ML model selection and training, ML model evaluation and fine-tuning.

Data Collection. Data gathering was a very important part of the work. The ESCO website [2] was the main place where the data was found. This platform offers a full grid of jobs and corresponding skills, which was used as the basis for training the ML model. ESCO is generally known as a reference for jobs and schooling in the European Union: it is like a dictionary for the European labor market. As the number of people who use ESCO grows, so does the number of ways the classification can be used. But not all users need the fine-grained information that ESCO skills or jobs provide. Some people like smaller, combined files that are easier to work with for their own reasons. To meet these needs, the European Commission has tried to make the ESCO dataset easier to understand by giving more active examples of how ESCO ideas can be linked together and used at more aggregated levels. The Commission has made matrix tables that link The International Standard of Occupations ISCO-08 work groups [18] to ESCO skills organizational groups. These tables, which come from the most thorough level of the ESCO classification, show what percentage of ISCO-08 work groups have ESCO skills. This method was very helpful in gathering data for our study: it made it possible to pull out relevant and doable data for the recommender system.

A Google Form poll was the second source of information. The goal of this survey was to get people's opinions about their skills and job choices: see Fig. 3. The answers were then put into a CSV file, thus building a complex model's testing data set. This two-source method made sure that the dataset was strong and varied, which made it easier for the model to make accurate recommendations. The fact that we used both the ESCO skills matrix table and other data assures that the job descriptions match also the local Romanian labor market status, as all our Google Form poll's respondents were

Romanians. All jobs/ occupations were described in various precents (between 0 and 1) by the following skills: handling and moving, information skills, working with computers, constructing, management skills, working with machinery and specialized equipment, assisting and caring, communication, collaboration, and creativity. If several respondents declared to have a certain profession (e.g., accountant), then an average of their perceived need for a certain skill was made.

occupation	handlingAndMoving	informationSkills	workingWithComputers	constructing	managementSkills	workingWithMachineryAndSpecialisedEquipment	assistingAndCaring	communicationCollaborationAndCreativity
Accountants	0.8	0.9	0.9	0.8	0.9	0	1	0.9
Accountants	0.7	0.8	0.9	0.8	0.7	0.7	0.9	1
Accountants	1	0.8	0.9	0.8	0.8	0.7	0.6	0.6
Advertising and marketing	0.9	0.9	1	0.8	0.7	0.5	1	1
Advertising and marketing	1	1	1	1	1	1	1	1
Advertising and marketing	0.7	0.6	0.8	0.7	0.7	0.7	0.8	0.7
Agricultural and forestry p	0.4	0.2	0.4	0.2	0.4	0	0.8	1
Applications programmers	0.5	0.5	0.6	0.5	0.6	0.5	0.7	0.7
Applications programmers	0.8	0.8	1	0.7	0.7	1	0.9	0.7
Applications programmers	0.7	0.7	0.9	0.8	0.9	0.7	0.9	0.9
Bartenders	0.8	0.8	0.7	0.8	0.7	0.7	0.8	0.8
Building architects	0.7	0.8	0.8	0.9	0.8	0.7	0.9	1
Business services and adm	0.8	0.8	0.8	0.9	0.7	0.8	0.7	0.8
Chemical engineers	0.9	1	1	0.9	1	0.8	0.7	0.8
Coding, proof-reading and	0.9	0.9	0.8	0.9	1	0.7	1	1
Computer network and sy	1	1	0.8	0.8	0.9	0.8	0.9	1
Computer network and sy	0.8	0.9	0.7	0.7	0.8	0.6	0.8	0.8
Computer network and sy	0.7	0.7	0.6	0.5	0.7	0.5	0.9	0.6
Computer network and sy	0.8	0.8	0.9	0.8	0.8	0.9	0.7	0.7
Data entry clerks	0.8	0.8	0.6	0.7	0.9	0.3	1	0.9

Fig. 3. Snippet of the Matrix Job-Needed Skills of CareProfSys Recommender System.

Data Preprocessing and Cleaning. Data preprocessing and cleaning is a crucial step in any data-driven project. For this research, both the training and testing datasets were cleaned using Excel. This process involved several steps: (1) any irrelevant or redundant information was removed to ensure the model was trained only on pertinent data; (2) missing or incomplete data was addressed to prevent any potential bias or inaccuracies in the model's recommendations; (3) the data was formatted appropriately to ensure compatibility with the ML algorithm. This meticulous cleaning process was essential in ensuring the integrity of the data and, by extension, the reliability of the model's recommendations.

Feature Extraction and Engineering. During feature extraction and engineering, the variables that the model would use to make its suggestions had to be found and worked on. In this case, the skills that go with each job were the variables, e.g., skill levels for working with computers, constructing, management, working with machinery and specialized equipment, for assisting and caring, for communication collaboration and creativity. These were taken from the datasets and then put on the same size to make sure they were all the same. Normalization is an important step because it makes sure that all features add the same amount to the distance calculation in the ML algorithm.

Model Selection and Training. The K-Nearest Neighbors (KNN) algorithm was picked for this study because it is good at making similarity-based recommendations.

KNN is an instance-based learning algorithm that predicts the values of new instances based on how close their features are to the ones already known. In this case, it was used to suggest jobs based on how close the skills were. The model was trained with data from the ESCO page about jobs and skills and tested with the data provided by our poll. The number of neighbors was set to eight, so the model could suggest up to eight jobs that are most like the user's skills. This parameter was picked so that users would get a variety of job suggestions without having too many to choose from [19].

Model Evaluation and Fine-tuning. Model evaluation and fine-tuning is a fundamental part of building any maker discovering model. For this study, the screening dataset from our Google survey was used to analyze as well as modify the design. The design's success was evaluated by how well it could discover jobs that fit the preferences and abilities of the people that filled out the form. This indicated contrasting the model's tips with the jobs that the interviewees desired. To make the design of the job profile much better, changes were made to its features, such as the number of next-door neighbors. This process of evaluating and tweaking the model repeatedly ensured that it was as accurate as well as dependable as feasible.

3.3 Validation of CareProfSys System in Real-Life Settings

To validate the recommender system, we made a 2-weeks experiment with 27 high school students from Romanian technological schools, who participated at a summer school organized by University POLITEHNICA of Bucharest. The purpose of the experiment was to investigate whether the recommended jobs are considered appealing to young people and whether they see themselves practicing those professions. The students had the opportunity to test the VR scenarios of the recommended jobs, thus simulating some basic activities from a possible future working day: see Fig. 4. Almost half (51.9%) of the participants have used emergent technologies like VR before. All the students were offered the same explanations and the same testing conditions.

Some interesting experiment data are listed below: (1) 63% of the students have not done anything related to the activities of the recommended occupation before, especially the ones who were advised by CareProfSys to become computer networking specialists or civil engineers; nevertheless, just 7.4% of them found the tasks specific to that profession to be difficult and very difficult; (2) all participants enjoyed the tasks in VR for the recommended profession (70.4% of them very much); (3) 59.3% of participants understood what he/ she should do in the future profession; (4) 85.2% of participants saw the VR scenarios as learning experiences (not just entertainment); (5) 88.9% of students considered VR scenarios to be helpful for describing professional occupations to young people; (6) 92.6% of students considered the idea of CareProfSys Job Recommender to be useful, because they needed career guidance and explanations about the activities they would do whether choosing a certain occupation. Most of the participants were happy to discover the recommended professions and saw themselves practicing those jobs in the future.

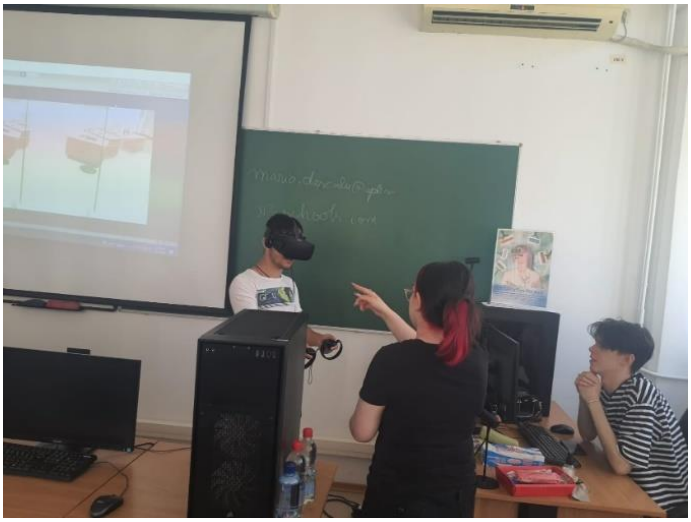

Fig. 4. Validation experiment of CareProfSys Recommender System.

Some validity threats for the positive response we obtained were: the fact that students were attracted by the gamification feature of the VR scenes, not by the activities specific to a certain profession, the fact that they wanted to be nice with us, as they were our guests and the fact that not all the recommended jobs has VR scenarios.

4 Conclusions

CareProfSys demonstrates how ML can be used to fix problems in real life: to implement a recommender system that can truly help young people discover their suitable jobs. The fact that CareProfSys does not offer only a textual description of the recommended profession, but a VR immersive experience, makes the recommender system to be especially useful for young people, a fact demonstrated by our experimental data. In the future, the hybrid recommended algorithm will be optimized by adding an ontological inference mechanism, the targeted ontology being the one reflecting the Romanian Classification of Occupations [20].

Acknowledgment

This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS–UEFISCDI, project number TE 151 from 14/06/2022, within PNCDI III: “Smart Career Profiler based on a Semantic Data Fusion Framework.”

References

1. Stanica, I.C., Hainagiu, S.M., Neagu, S., Litoiu, N., Dascalu, M.I.: How to Choose One's Career? A Proposal for A Smart Career Profiler System to Improve Practices from Romanian Educational Institutions. In: 15th annual International Conference of Education, Research, and Innovation Proceedings (ICERI 2022), pp. 7423-7432. IATED, Seville, Spain (2022)
2. ESCO: Skills-Occupations Matrix Tables, https://esco.ec.europa.eu/en/about-esco/publications/publication/skills-occupations-matrix-tables , last accessed 2023/06/20
3. Turing: How Does Collaborative Filtering Work in Recommender Systems?, https://www.turing.com/kb/collaborative-filtering-in-recommender-system#user-item-interaction-matrix, last accessed 2022/10/22
4. Iterators: Collaborative Filtering In Recommender Systems: Learn All You Need To Know, https://www.iteratorshq.com/blog/collaborative-filtering-in-recommender-systems/, last accessed 2023/10/28
5. Career Explorer, https://www.careerexplorer.com/assessments/, last accessed 2023/05/27
6. *NET Interest Profiler, https://www.mynextmove.org/explore/ip, last accessed 2023/06/25
7. Pymetrics, https://www.pymetrics.ai, last accessed 2023/06/25
8. Job Scan, https://www.jobscan.co, last accessed 2023/06/25
9. My Next Move, https://www.mynextmove.org, last accessed 2023/06/25
10. Sampaio, A.Z., Martins, O.P.: The application of virtual reality technology in the construction of bridge: The cantilever and incremental launching methods. Autom. Constr. 37, 58–67 (2014)
11. Tay, Y.X., McNulty, J.P.: Radiography education in 2022 and beyond - Writing the history of the present: A narrative review. Radiography 29(2), 391-397 (2023)
12. Harknett, J., Whitworth, M., Rust, D., Krokos, M., Kearl, M., Tibaldi, A., Bonali, F.L., Van Wyk de Vries, B., Antoniou, V., Nomikou, P., Reitano, D., Falsaperla, S., Vitello, F., Becciani, U.: The use of immersive virtual reality for teaching fieldwork skills in complex structural terrains. Journal of Structural Geology 163, 104681 (2022)
13. Angular, https://angular.io/ , last accessed 2023/06/25
14. Flask, https://pythonbasics.org/what-is-flask-python/, last accessed 2023/06/25
15. Python, https://www.python.org/doc/essays/blurb/, last accessed 2023/06/25
16. WebG , https://developer.mozilla.org/en-US/docs/Web/API/WebGL_API, last accessed 2023/06/25
17. MongoDB, https://www.mongodb.com/, last accessed 2023/06/25
18. ISCO-08, https://www.ilo.org/public/english/bureau/stat/isco/isco08/, last accessed 2022/10/17
19. Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K. (2003): KNN Model-Based Approach in Classification. In: Meersman, R., Tari, Z., Schmidt, D.C. (eds) On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM 2003. Lecture Notes in Computer Science, vol 2888. Springer, Berlin, Heidelberg (2003)
20. Dascalu, M.I., Bodea, C.N., Nemoianu, I.V., Hang, A., Puskás, I.F., Stanica, I.C., Dascalu, M.: CareProfSys – An Ontology for Career Development in Engineering Designed for the Romanian Job Market. Rev. Roum. Sci. Techn.– Électrotechn. et Énerg. (RRST-EE) 68 (2), 212–217 (2022)