Advanced Engineering Informatics 52 (2022) 101508

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

An AI-based open recommender system for personalized labor market driven education

Mohammadreza Tavakoli [a], Abdolali Faraji [a], Jarno Vrolijk [b], Mohammadreza Molavi [c], Stefan T. Mol [b,*], Gábor Kismihók [a]

[a] Leibniz Information Centre for Science and Technology (TIB), Germany
[b] University of Amsterdam, The Netherlands
[c] Amirkabir University of Technology, Iran

ARTICLE INFO

Keywords:
Recommender systems
Open educational resources
Educational data mining

ABSTRACT

Attaining those skills that match labor market demand is getting increasingly complicated, not in the last place in engineering education, as prerequisite knowledge, skills, and abilities are evolving dynamically through an uncontrollable and seemingly unpredictable process. Anticipating and addressing such dynamism is a fundamental challenge to twenty-first century education. The burgeoning availability of data, not only on the demand side but also on the supply side (in the form of open educational resources) coupled with smart technologies, may provide a fertile ground for addressing this challenge. In this paper, we propose a novel, Artificial Intelligence (AI) driven approach to the development of an open, personalized, and labor market oriented learning recommender system, called *eDoer*. We discuss the complete system development cycle starting with a systematic user requirements gathering, and followed by system design, implementation, and validation. Our recommender prototype (1) derives the skill requirements for particular occupations through an analysis of online job vacancy announcements; (2) decomposes skills into learning topics; (3) collects a variety of open online educational resources that address those topics; (4) checks the quality of those resources and topic relevance with three intelligent prediction models; (5) helps learners to set their learning goals towards their desired job-related skills; (6) recommends personalized learning pathways and learning content based on individual learning goals; and (7) provides assessment services for learners to monitor their progress towards their desired learning objectives. Accordingly, we created a learning dashboard focusing on three *Data Science* related jobs and conducted an initial validation of eDoer through a randomized experiment. Controlling for the effects of prior knowledge as assessed by means of a pretest, the randomized experiment provided tentative support for the hypothesis that learners who engaged with personal recommendations provided by eDoer to acquire knowledge of basic statistics, attained higher scores on the posttest than those who did not. The hypothesis that learners who received personalized content in terms of format, length, level of detail, and content type, would achieve higher scores than those receiving non-personalized content was not supported.

"Ensure Inclusive and Equitable Quality Education and Promote Life-long Learning Opportunities for All" - United Nations Sustainable Development Goal 4 [1].

1. Introduction

With the clock ticking on the United Nations' Sustainable Development Goal pertaining to quality education, the time is ripe to develop cost effective, scalable, and sustainable means to match the exponentially growing array of open educational resources to (the needs of) individual learners, regardless of socio-economic status and/or demographic background. Indeed, where top-quality educational resources were once solely accessible to the privileged few who were enrolled in top-tier educational institutions (mostly) in developed nations, the growing trend of opening up such resources, together with technological developments that allow for the matching of content to learners on a massive scale, has created opportunities to distribute and disseminate

* Corresponding author.
E-mail addresses: reza.tavakoli@tib.eu (M. Tavakoli), abdolali.faraji@tib.eu (A. Faraji), j.vrolijk@uva.nl (J. Vrolijk), mr.molavi@aut.ac.ir (M. Molavi), s.t.mol@uva.nl (S.T. Mol), gabor.kismihok@tib.eu (G. Kismihók).

https://doi.org/10.1016/j.aei.2021.101508
Received 30 April 2021; Received in revised form 10 November 2021; Accepted 19 December 2021
Available online 24 February 2022

such educational resources more equitably, inclusively, and effectively, to all those who seek them.

Although altruism and the "feel-good factor" have been identified as some of the main drivers of the movement to open up educational resources [2], to date the word *open* has remained more of a legal designation, than a harnessed potential. There are benefits associated with tapping into the vast array of Open Educational Resources (OER) that go beyond just making them accessible to people who may otherwise not be able to access education. First, in light of the burgeoning amount of publicly available textual data [3], there are opportunities for more explicitly mapping educational content to the demands of the labor market, therewith enhancing learners' motivation, learning effectiveness, and employability. Indeed, to date, efforts at personalizing educational content to learners is often backward-looking (i.e., where learners came from) as opposed to forward-looking (where they are going). Second, greater and greater demands are being placed on teachers, not only in terms of the ICT (Information and Communications Technology) heavy teaching methods they need to master, but also in terms of increasing student numbers and courses they may have to teach. As we shall illustrate later, the ability, on the part of students, to automatically identify and be recommended OERs based on where they stand and where they are going may complement traditional courses and may ultimately serve to make teachers' workloads more manageable. Third, many educational curricula crush student self-directed learning, proactivity, sense of control, and autonomy by dictating what is to be learned and when it is to be learned, without providing learners with a sense of the bigger picture, or why they are learning what they are having to learn. The information asymmetry that this entails, means that all too often students are just passive receivers of education, as opposed to them taking guided decisions and expending motivated effort towards shaping their own future. It is against this backdrop that we started working on designing and constructing a vehicle that can connect learners to the educational contents that they seek and/or need regardless of their geographic location, demographic characteristics, and/or formal educational qualifications.

Recent decades have seen educational environments changing dramatically in response to the increasing demand for online personalized learning [4]. There is a growing need for online personalized educational services because of (1) the rapid evolution in both the quantity and quality of skills demand [5–7]; (2) the gap between knowledge (and skills) that job markets require and the training that formal educational programs offer [8–11]; and (3) the global challenges for work and education due to the emergence of the COVID-19 pandemic [12].

Consequently, we are facing exponential growth in educational resources (such as Online and Open Educational Resources) that are being produced and disseminated on an unprecedented scale, and published in different contexts (e.g., location, language, discipline, expertise level, and format) [13,14]. However, the heterogeneity and (lack of) targeted distribution of these educational contents leads to a number of problems for learners that limit their usefulness. First, learners may not understand which components they need to learn to fulfill skill (or knowledge) requirements [15] let alone the skills demanded by the labor market. Second, even if they knew what it was that they were seeking to learn, learners are unlikely to be able to distinguish between high quality and low quality educational resources. In sum, confronted with an abundance of learning materials, learners may be overwhelmed and will unlikely be able to plot and follow their own effective learning path without directional guidance in the form of personalized educational recommendations.

According to the state of the art and our requirement gathering, there is a need for intelligent systems that help learners to (1) be up-to-date about the required skills for their target (current or future) jobs; (2) be informed about the components (topics) that they need to learn for each skill; (3) build their own learning path towards the required skills; (4) find personalized learning materials according to their path; and finally (5) assess (changes in) prerequisite knowledge for requisite skills.

A number of occupational taxonomies (e.g., ESCO and O*NET) exist that may be leveraged to provide information about occupations. However, most of these taxonomies are updated through a largely manual process, meaning they are time-consuming and expensive to construct, and also susceptible to being outdated [16]. Alternatively, text-based algorithms can be developed to extract those topics that are on the one hand manifest in corpora of job vacancies, and on the other, covered by existing educational materials in an effort to help learners to build their learning path [13–15,17]. To offer personalized educational services (e.g recommendation and search services), we thus need to (1) extract properties of educational materials (e.g., quality and metadata [18,19]), (2) detect preferences of learners (e.g., preferences regarding format and the time investment associated with mastering the educational content [4,12]), and (3) match between the resources and the learners [20,21]. However, previous efforts to build such educational systems by drawing on the rapidly growing amount of online and open educational resources [22,23] revealed the lack of high-quality OER metadata, and quality control [18]. These issues seriously curtail the accessibility of OERs, an issue that may be tackled through the deployment of high-quality OER search and recommendation services [20,21]. Our aim, therefore, besides contributing to the resolution of OER metadata quality issues by leveraging the work of [12], is to generate metadata (e.g., technical quality of video/sound/text, and how relevant the content is to a target topic) by having learners interact with our OER recommendation system (i.e, through point-of-experience surveys that solicit quality feedback).

In this paper, we set out to address the aforementioned challenges and present a software prototype (with an initial focus on data science related jobs) to provide personalized, open educational content recommendation in an effort to help learners master their target skills by:

- Defining those topics that need to be learned to acquire a skill
- Empowering learners to construct their own learning trajectories based on labor market information and OERs
- Recommending personalized educational resources based upon our automatic quality control models
- Providing a two-layer assessment that evaluates the level of the learner both with regard to target skills and topics.

We end this article by describing our effort to validate our prototype by conducting a randomized experiment in which 175 paid participants were asked to learn about *Basic statistics for engineers*. The specific objectives of this study were (i) to examine whether learners who engaged with our system acquired more knowledge than learners who did not, and (ii) to examine, among those users who engaged with the system, whether learners who received personalized content recommendations acquired more knowledge than those who did not.

This paper is organized as follows: Section 2 depicts the state-of-the-art regarding the different components of our proposed system, followed by Section 3, in which we explain the process of constructing our proposed system including the requirement gathering/analysis step, matching jobs with their required skills, decomposing skills into meaningful educational components (i.e., topics), collection and quality control of educational resources, building our recommendation system, and implementing our proposed learning dashboard. Section 4 presents our hypotheses, and discusses the methods and outcomes of the validation study into our prototype. Subsequently, in Section 5 we summarize the main findings pertaining to to our research questions, objectives, and hypotheses before discussing the implications, limitations, and future directions. In Section 6 we conclude the paper.

2. Related work

2.1. Automatic extraction of online educational content properties

In order to build their own learning trajectory [17], it is imperative for learners to have insight into the specific topics and properties that are addressed by particular Online Educational Resources. To tackle this issue, some studies have developed semantic [13,14,24] and machine learning based [15,25,26] methods to extract topics and other pertinent metadata from educational resources. For instance, summarized content, derived by extracting keywords from educational resources [27], may be linked to other relevant contents such as *Wikipedia* [28]. Other studies have focused on extracting the type of educational content. For example, [29] applied deep neural architectures to extract the video types (e.g., PowerPoint presentation, code writing, or the instructor talking to the learners) from educational videos. Although further work is sorely needed in this area [15], these studies illustrate the viability of automatically extracting properties from educational resources that may then be used as a basis for matching and recommendation.

2.2. Quality evaluation of online educational content

The sheer amount and heterogeneity of online and open educational contents that are created and uploaded to the internet on a daily basis, places limitations on that which can be achieved with manual quality control. This is problematic, because high-quality metadata are a mandatory prerequisite of any data driven (not only educational) service [30]. In light of this, some studies have attempted to define metrics (e.g., completeness and consistency of metadata [31], provenance, and accuracy of metadata [32]) to assess OER metadata quality. [33], based on existing developments in quality models in e-learning, semantic-based methods, and NLP techniques, developed a quality assessment framework. Building on this line of work, [18] developed a metadata scoring method, which can be applied to the definition of a model to predict OER quality. Also, [19] proposed a quality evaluation model for open educational videos: their quality was effectively predicted on the basis of video transcripts, popularity metrics (i.e., ratings, likes, dislikes), and length. Despite the progress in this area, the fast development of disparate online and open educational resources calls for researchers to continue to improve the methods that underpin quality evaluation models [18].

2.3. Recommending open online educational content

The literature on high-quality OER recommender systems is very limited [21]. However, in recent years, some studies have advanced recommendation algorithms based on ontologies, linked data, and open-source Resource Development Framework (RDF) data to leverage semantic content [21,23]. For example, [22] defined an ontology for learners, learning objects, and their environments to provide adaptive recommendations, based on similarities between object properties. In a different vein, [20], using existing ontologies, examined the *Cold Start problem [34]* in the realm of micro OERs, by defining rules, based on recommended sequences of learning objects. Finally, [4,12] built an OER recommendation system to help learners achieve skill-based learning objectives using (1) a text mining approach to extract skills from online job vacancies, and (2) a gradient descent algorithm to predict user preferences based on their ratings of previously recommended educational resources.

2.4. Content recommendation for further education

Due to the dramatic evolution of skill requirements in recent decades, matching processes between skill demand and supply on the labor market are getting more and more complicated as skills dynamically change through an uncontrollable and seemingly unpredictable process [6,7]. Employees' staying abreast of these changing labor market demands, and acquiring the concomitant skills is a prerequisite for sustainable employability [6,35]. However, when it comes to the offerings of Educational institutions these often lag behind, increasingly resulting in a gap between the skills (and associated learning content) that educational institutions impart in their students, and the current skills that job markets demand [9,11,36]. To mitigate this mismatch, understanding labor markets dynamics and information is critical [4]. On the one hand, this understanding requires the decomposition of jobs into skills and other relevant building blocks. Recent efforts in this area have applied text mining/machine learning techniques [3,37] and semantic-based [38] methods to text data obtained from online vacancies. On the other hand, the extracted skills from the demand side must go through a matching process, by collecting related high-quality educational resources which may address the learners' needs towards achieving those skills, and using recommendation methods (e.g., Content-based and Collaborative filtering) to provide the most relevant content for learners to help them acquire those skills efficiently [4,12]. According to the extant literature, the area of labor market based education requires more research and development to facilitate the provision of more personalized and up-to-date services to individual learners.

2.5. Research questions and objectives

Based on the state of the art, it is clear that (1) online educational services that offer open and online educational resources to learners may address the pressing need for inclusive, equitable, and effective online education, while at the same time becoming ever more viable as technology is developing. Therefore (2) the time is ripe to focus greater research attention on personalized high-quality educational resource recommendations. (3) This is especially important for learners in further education, who need help to build their own learning trajectories towards their desired jobs, thus (4) offering learners those existing educational contents that they need to learn in order to acquire the skills which are demanded by the labor market, can potentially enhance learners' autonomy, and motivation, thereby enhancing the (effectiveness of the) individual learning experience. In light of these points, the main research questions of this paper are:

- *RQ1*. What are the key learner requirements (e.g., vis-a-vis type of content, level of detail, and assessment) for labor market driven personalized education?
- *RQ2*. How may learners be aided in constructing their own learning trajectories based on labor market information and open/online educational resources?
- *RQ3*. What models provide adequate and accurate predictions about the quality of online educational content?
- *RQ4*. How can we combine job/skill decomposition, quality prediction of educational resources, and feedback to learners to develop a personalized open educational content recommender system for learners in further education?

3. Method

3.1. Requirement analysis

First, we collected relevant stakeholder requirements to further define our objectives and guide our investigation. For this exercise, we built an initial and bare-bones OER recommender prototype so as to be able to showcase our approach to key stakeholders. Through qualitative interviews, this prototype was evaluated by 23 subject matter experts (e.g., university instructors and Ph.D. students) with significant experience in both industry and learning/teaching [4,12].

Based on their feedback, we designed a questionnaire[1] to capture the needs of those stakeholder groups that we expected to be potentially important beneficiaries of our learning recommender system. We identified the following stakeholder groups (personas)[2]:

- **Group1.** Recipients (e.g., Learners, Researchers, Students)
- **Group2.** Deliverers (e.g., Professors, Lecturers, Study Counselors)
- **Group3.** Facilitators (e.g., Managers, Educational Support Staffs)

We obtained 13 potential user requirements from the initial qualitative interviews, which we then presented to survey participants (see Table 1), asking the latter to rate those in terms of their importance and frequency of use. Since in this study we focus solely on the learner perspective, the following subsections showcase the most important outcomes and lessons learned from **Group1** members.

3.1.1. Personal information

Altogether 47 learning recipients (*Group 1*) from 10 countries completed our questionnaire and returned usable data. Of these **Group1** participants, 43.2% were female, 51.3% were male, and 5.5% did not provide any information on their gender. Of the participants, 12.8% had completed High-school or lower, 14.9% had a Bachelor, 36.2% had a Master, 34% had a Ph.D., and 2.1% had completed other educational degrees or qualifications.

3.1.2. Current skill progression towards desired occupations

Survey participants' reported informing themselves about skill demands in the following ways: 86.5% while performing their everyday tasks, 62.2% through reading related papers or news, 54.1% by inquiring with their supervisors, and 40.5% through job vacancy announcements of positions they apply to. Moreover, they mentioned courses (83.8%), educational videos (78.4%), books (72.9%), and Web pages/documents (64.9%) as dominant resources they used to develop themselves towards skills required by employers. Finally, with respect to open learning content for their self-development, participants bemoaned (1) the lack of personalization, (2) the identification/localization of high-quality learning content, and (3) the time-consuming search process, as the most pressing problems.

3.1.3. Importance and frequency of use of the potential requirements

Participants rated the importance (1: Not at all important - 5: Very important), and frequency (1: Never - 5: Daily) of usage for each potential user requirement. Once data collection was complete, we calculated the average of their ratings for each of the requirements and normalized the average rates using *Min–Max Normalization* as (1) in which we replaced the *values* with the *average rates*. Table 1 shows the potential user requirements, normalized average importance ratings, normalized average frequency ratings, and the composite rate (multiplication of the normalized importance and frequency rates) which have been sorted based on the composite rates.

$$Normalized\_value = \frac{Value - Minimum_{values}}{Maximum_{values} - Minimum_{values}} \quad (1)$$

[1] The questionnaire is available on: https://tib.eu/umfragen/index.php/survey/index/sid/977178/newtest/Y/lang/en.

[2] It should be mentioned that we allowed participants to answer our questionnaire from the perspective of multiple personae. This was important, as a single person can fulfill different roles in a learning process (e.g., a person can be a lecturer and manager at the same time).

3.1.4. Lessons learned

By analyzing participants' ratings regarding these potential user requirements, we prioritized and constructed the following services for learners:

- ***Service_1*: Personalized Search.** $Req1$ and $Req2$ (Table 1), clearly received the highest ratings among all requirements. Therefore implementing an educational resources search service, which provides accurate and high-quality search results to address individual learning needs, became one of our top priorities. Clearly, the personalization and the content-quality of the results of such a service are critical as demonstrated by 3.1.2, where learners pointed to the *lack of personalization* and *problems in identifying high-quality learning content* as two of the most important barriers to using open/free educational resources. Hence we focused on the context of the learners (e.g., job, skill-set, expertise level, language), and their learning preferences (e.g., their preferred format (e.g video or web pages)).
- ***Service_2*: Goal-driven Learning Content Recommendations.** According to $Req2$, $Req3$, and $Req5$, learners desire a service that helps them (1) explicate their learning objectives, (2) find suitable learning pathways that fit to their context (preferences), and (3) receive the most relevant and highest-quality learning resources needed to meet their learning objectives.
- ***Service_3*: Elucidating Job Skill Requirements.** Based on $Req4$ and $Req6$, the need can be observed to match jobs and the skills that are required to be effective in those jobs. This should be accompanied by visualization, which helps inform users about those skills they need to acquire. Based on this information one can set learning targets and obtain (and ultimately learn) relevant learning content.
- ***Service_4*: Learning Progress Monitoring.** Learners also expressed a strong interest in monitoring their progress towards their learning goals ($Req8$). Accordingly, we found it essential to provide an assessment service, which would help users to test the knowledge they set out to acquire. Additionally, we decided to provide further insights (through numbers, charts, etc.) about users' progress towards each of their learning goals.

3.2. Labor Market Intelligence

In order to match jobs to their skill requirements ($Req4$ and $Req6$), we deployed a Labor Market Intelligence (LMI) component to capture up-to-date skill requirements for jobs relevant to this study.

In an initial effort demonstrate the applicability of our system, we decided to focus on *Data Science* related jobs. We did so because these jobs are both in high demand and particularly prone to change. We selected three associated jobs: *Data Scientist, Data Analyst,* and *Business Analyst*. Subsequently, we used a sample data-set of English job vacancies from *Monster.com*,[3] which included 21,937 vacancies and their related skills.

Subsequently, we calculated the rate of occurrence for each of the skills in the target jobs and set the importance of the skills in each job based on this occurrence rate. We used this importance rate to sort the skills that learners need to learn. Based on this process, the following six skills were selected to represent our target jobs as they achieved the highest importance rates across our target jobs: *Python programming, R programming, Statistics, Machine learning, Data Visualization, and Text mining*.

[3] The data-set is available on: https://www.kaggle.com/PromptCloudHQ/us-jobs-on-monstercom/version/1.

Table 1
Average importance and frequency ratings for potential user requirements.

Requirement	Importance rate [0–1]	Frequency rate [0–1]	Composite rate [0–1]
*Req*1. Finding learning content about a problem I am working on at the moment	1.00	1.00	1.00
*Req*2. Identifying high-quality content which fulfills my learning needs	0.81	0.68	0.55
*Req*3. Knowing where to start learning when I need a new skill for my studies/job	0.75	0.38	0.29
*Req*4. Identifying which skills are required for my current/future job	0.70	0.36	0.25
*Req*5. Defining my own goals towards jobs I find attractive	0.53	0.29	0.15
*Req*6. Identifying which skills are required for my degree	0.40	0.27	0.11
*Req*7. Finding out how I can improve my skillset in order to qualify for my desired job	0.58	0.18	0.10
*Req*8. Monitoring my learning progress towards desired skills	0.23	0.24	0.06
*Req*9. Making sure that my learning objectives meet job requirements	0.40	0.11	0.04
*Req*10. Identifying which skills are the most important ones in terms of contributing to expected salary	0.05	0.07	0.004
*Req*11. Visualizing potential skill targets	0.05	0.05	0.003
*Req*12. Identifying which jobs I can fulfill with my skillset	0.15	0.00	0.00
*Req*13. Visualizing the structure of the content that I need to master to achieve my skill targets	0.00	0.04	0.00

Table 2
Collected resources for each skill.

Skills	Number of collected playlists	Number of covered educational videos	Number of topics
Python programming	8	502	26
R programming	4	185	12
Statistics	9	621	27
Machine learning	9	472	35
Data visualizing	8	257	14
Text mining	6	194	18

3.3. Educational topic detection for selected skills

In order to recommend open learning content for the selected skills (*Req*2 and *Req*3), we needed to decompose each skill into meaningful learning *Topics*. Therefore, we extracted learning topics for these six skills by applying *Latent Dirichlet Allocation* (*LDA* [39]) to the transcripts of existing educational materials. Specifically, we used the method proposed by [15] to extract learning topics and determine the degree to which those topics were reflected in each educational resource. Finally, we asked three experts to prioritize each of the extracted topics with an eye on skill development. Table 2 shows the number of collected playlists (each of which comprises the educational resources per skill), the number of covered educational videos, and the final number of extracted topics for each skill. It should be mentioned that some of the topics were part of more than one skill (e.g., *Linear Regression* was a topic of both *Machine Learning* and *Statistics* skills)

3.4. Incorporation of educational content

In this section, we describe how relevant high-quality open educational resources were collected, filtered, and labeled (*Req*1 and *Req*2). We also depict how assessments were connected to the final set of educational resources included in our recommender (*Req*8).

3.4.1. Collection of online educational resources

To collect open educational content for the six skills and their topics, we performed a search on *Google* and *Youtube*[4] using the concatenation of the skill and the topic (e.g., "Python programming Conditions") as the search keywords. We collected 3,228 educational resources[5] which included 2,514 educational videos and 724 text-based resources (e.g., web pages, lecture notes, and book chapters). For each resource, we collected the following fields based on the available fields for online and open educational resources and the fields we needed to apply our automatic models[6]:

- *Source*. Records the original location of the content.
- *Format*. The format (e.g., Video, Web page, or Book chapter) of the content. This was set based on the source and file extension of the resources. For example, this field was set to *Video* for the resources from *Youtube*.
- *Title*. Records the title of the content.
- *Description*. Records the description of the content.
- *Transcript*. Records the transcription of the content. This field was set based on the transcript of the videos, and the content of the web pages, and book chapters.
- *Rating*. User ratings of the content. This field was calculated differently (e.g., based on 5 point rating scales or likes and dislikes) in the different sources. Therefore, we normalized the ratings for each of the resources.
- *Length*. This field shows the content length (in seconds only for videos).
- *View Count*. Total number of times that the educational content had been viewed by users.

3.4.2. Filtering based on quality and relevance

To provide high-quality educational content, which was one of the key outputs of our requirement analysis step (*Req*2), we applied the following filtering procedure on the collected OERs and other available educational resources:

- **Topic-based filtering**. In order to remove educational content that did not fit the search keywords detailed in the previous section, we used the output of our topic models that was described in Section 3.3. Specifically, we extracted the target topic of each educational resource using our topic models, and removed those resources for which the extracted target topic did not match its search keywords. For instance, if a video was the result of the search keywords "Machine Learning Linear Regression", but our model detected its focus as "Support Vector Machine", we removed it from our resource list. This step resulted in the removal of a total of 1,116 resources (906 of which were video- and 210 of which were textual resources)
- **Metadata-based filtering**. Previously, [18,40] showed that the metadata quality of OERs is indicative of their content quality. Based on this finding, we created a binary classifier to sort educational resources into a *high-quality* and a *low-quality* group. By applying their machine learning model, educational resources with a predicted low-quality content (a total of 727 resources of which 621 were video-and 106 were textual resources) were removed from our educational content collection.

[4] Using Pafy python-youtube library: https://pypi.org/project/pafy/.
[5] This is a new data-set and is different from the one we used for the topic detection step.
[6] It should be mentioned that some resources in our data-set did not include all the mentioned fields.

Table 3
Number of resources which passed through our filtering steps.

Skills	Number of educational resources	Avg number of resources per topic
Python programming	124	4.77
R programming	49	4.08
Statistics	209	7.74
Machine learning	263	7.51
Data visualizing	100	7.14
Text mining	120	6.67

- **Quality-based filtering**. In our last filtering step, we checked whether OERs and other available educational resources fit the description of the target learning goal of the content (based on the *Wikipedia* page of the search keyword we used to collect the content), and the level of prior learners' satisfaction in terms of content ratings and view counts. This was accomplished through the quality prediction model proposed by [19]. This model leverages the similarity between the transcription of educational resources and the description of their target topics (from *Wikipedia*) in addition to their popularity features (e.g., rating and view count) to determine quality. To apply the model on our data-set, we rebuilt their proposed prediction model based on the features that existed in all of our collected resources (i.e., Transcript, Rating, and View Count) which led to 79.2% of the F1-score on their published data-set. As a result of this step, a total of 631 (547 video- and 84 textual) resources were removed from our collection.

Through the application of the aforementioned filters, we distilled 764 high-quality (440 video- and 324 textual) OERs and other available educational resources, covering all topics 3.3 in our six target skills (see Table 3). It should be noted that the number of educational resources for each topic ranged between 3 and 10, and that we had at least one video and one textual resource for each. Moreover, in our data-set, there were resources that addressed more than one topic (e.g., an educational video could cover both *Linear Regression* and *Gradient Descent*).

3.4.3. Educational resource labeling

To generate the personalized recommendations for the learners, we analyzed and labeled all of the educational resources that were retained. Some features such as *Source, Format, Transcript, Rating, and View Count* had already been extracted automatically (see 3.4.1). Additionally, for each skill, we asked two experts to review and label the resources (see below). As a result, the following features were collected for all filtered educational resources:

- *Length.* As we extracted the length of educational videos (in seconds), we asked experts to estimate how long it would take learners (in seconds) to scrutinize the text-based educational resources. Afterward, we grouped educational resources in such a way that we had groups with a similar number of resources, that we could describe to the learners easily. Therefore, we created 3 groups of *Short* < *10 min* (included 308 resources), *10 min* < *Medium* < *20 min* (included 225 resources), and *Long* > *20 min* (included 231 resources) resources.
- *Level of Detail.* This feature captures the level of detail in which a specific content addresses a target topic.[7] Experts assigned the following labels to the resources: *Low Detail, Medium Detail, High Detail.*
- *Learning Strategy.* We defined three learning strategies of *Theory-based, Example-based,* and *Mixed* (which includes both theory and example) based on [41], and asked experts to label resources accordingly.
- *Is a Classroom-Based Instruction.* This field is a Boolean value that captures whether the resource has been recorded as a university class or not.

3.4.4. Implementation of learning progress monitoring

To produce well-defined and relevant assessments, three experts generated and carefully reviewed multiple-choice questions (test items) for each topic. In this process, a question was selected to be added to our test items, when all reviewers found it appropriate to assess the knowledge of learners in the topic(s) that the question targeted. This resulted in a repository of topic-based and skill-based test items. In our prototype, we implemented two different types of assessment, each of them are generated dynamically, according to the individual progress of each learner:

- A **progress assessment** is a test which only contains test items related to topics. This test validates the progress of a user, when they transit between consequent topics within a skill. Learners can only start a new topic if they pass the assigned progress test of the prerequisite topic(s). In case a learner passes a topic associated with a target skill, the topic is marked as completed in all corresponding (related) skills listed in our recommender. For instance, if a learner passes the topic "Linear Regression Concept" when studying for "Machine Learning" skill, this topic will also be completed for the skill "Statistics", even though this skill is not among the skill targets of the learner. This method helps track individual development, by monitoring knowledge and skill proficiency levels across topics and skills, within and beyond individual learning objectives.
- A **skill assessment** can be interpreted as an assessment of skills (i.e., a topic aggregate), and can be used to provide feedback to learners about applying and combining acquired knowledge (topics) areas in relation to a specific skill. Therefore, these assessments include questions that cover all topics associated with a specific skill. Learners complete these assessments as soon as they have mastered the different components (topics) of a target skill.

Using progress and skill assessments, a learner can continuously evaluate their level of knowledge in a fine-, and coarse-grain manner.

3.5. Personalized open learning content recommendation

In this section, we demonstrate our proposed personalized recommendation system for learners to address $Req1$ and $Req2$.

3.5.1. Learner profile

Based on the features we collected for educational resources 3.4, we also defined features for the preferences of each and every learner. These features are described in Table 4. Based on possible feature values, we created a long-, and a short-term 15-dimensional preference vectors for each learner which included the following features: *Length-Short, Length-Medium, Length-Long, Detail-Low, Detail-Medium, Detail-High, Strategy-Theory, Strategy-Example, Strategy-Both, Class-based, Non-class-based, Content-Video, Content-Book Chapter, Content-Web Page, Content-Slide*. Each feature value in a vector shows how much (a float value from 0 - the lowest, to 1 - the highest) a learner prefers receiving learning resources with that feature. The long-term vector is used as the basis for our learning content recommendation. Therefore, the complete history of each learner's feedback (5-scale ratings for the recommended educational contents) until the recent updating period is taken into account. The short-term vector shows learners' feedback in the recent updating period (last one month) and it affects the long-term vector at the end of each updating period; therefore, the short-term vector is emptied at the starting point of each updating period and updated after each feedback from the learner. The long-term

[7] The topic can be a concept, formula, or an API.

Table 4
Preference features.

Feature	Possible values	Notes
Length	Short, Medium, or Long	Learner's preference about the length of educational resources
Detail	Low, Medium, or High	Learner's preference about the level of details in educational resources
Learning Strategy	Theory-only, Example-only, or Both	Learner's theoretical knowledge orientation
Classroom-based	Yes or No	Learner's preference about learning content originated from classrooms
Content Format	Video, Book, Web page, Slide	Learner's preference about learning content formats

vector helped us to capture the learners' preferences while using the recommended resources (it should be noted that the long-term vector is configured to place more weight on the recent ratings). We defined the *updating period* as a configurable period value (which could be set in our system), and set it to one month in this version of our prototype.

When a learner registers in our system for the first time, we ask questions regarding all preference features in order to populate the long-term preference vector. This is done by transforming the selected values into to the corresponding values (float number between 0 and (1) in our preference vectors. For instance, when a learner prefers *Long* content, the *Length-Long* feature is set to 1, while the Length-Short and Length-Medium features are set to 0. As another example, if a learner, selects 3 on a 5 point rating scale rating regarding the video contents, the Content-Video feature is set to 0.5.

When the learners complete a learning content, we consider their feedback, which is a 5-scale rating, to update their short-term profile. For instance, assume that after recommending two pieces of learning content with a *High* level of detail to a learner, and we receive the following feedback ratings: (1) 3 in a 5-scale rating (which means 0.5 out of 1 in our system), and (2) 5 out of 5 (which means 1 out of 1). As a consequence the *Detail-High* feature of the short-term vector is set to 0.75 (which means 4 in a 5-scale rating) for the learner.

At the end of each updating period (which was set to one month), we updated the long-term vector by calculating the average of the current long-term vector and the short term vector. This updating procedure detects changes in long-term individual learning preferences and results in more relevant content suggestions. It should be mentioned that the values of the long-term vector can be also viewed and directly edited by learners through their dashboard, in their profile settings.

3.5.2. Recommendation engine

To recommend learning content on a specific topic to a particular learner, first we retrieved all the resources (the ones that passed our filtering process) which focused on the topic. Afterward, we created the same 15-dimensional vector (with the same features as the preference vector) for each retrieved learning resource, as we did for the learners (see Section 3.5.1).[8] Finally, we calculated the *Dot Product* [42] of the learner's long-term preference vector together with the created vectors of each retrieved learning content. As a result, our system recommends the content with the highest *Dot Product* result.

3.6. Learning dashboard

In this section, we showcase our learning dashboard, called *eDoer*, that we implemented to provide our individualized learning services 3.1.4 to learners.[9] Fig. 1 illustrates how the different technical components of our recommender prototype interact with one another (and with the learner) to create the learner's personal learning experience 3.1.4. For the User Interface (UI) we incorporated responsive web design and design guidelines [43,44]. We provided learners with an interactive tutorial [45] at their first login, in order to familiarize them with the different functionalities of our learning dashboard.

3.6.1. Registration and goal setting

The registration path consists of three consecutive steps, each serving a different purpose: (1) In the first step we collect the necessary demographic information from new learners, including their name, email address, gender, and geographical location (country and city).[10] (2) In the second step, learners search for and select a target job. Subsequently (as an implementation of *Service_3* depicted in 3.1.4), we show the required skills for the selected job by using our *labor Market Intelligence* 3.2, and ask learners to select those skills they want to master. In addition, users can search and select complementary skills (not connected to their target job) and add them to their target skills manually.[11] (3) The third (and last) step consists of setting learning preferences by answering a number of questions (see Section 3.5.1), to further calibrate the learning content recommender algorithm for each particular learner.[12]

3.6.2. Personalized learning

To provide *Service_2* (see 3.1.4), a curriculum page was designed to structure and monitor the advancement of learners with respect to their target skills and related topics. Learners can visualize their personalized curriculum by selecting a skill. Once the skill is selected, the related list of topics are displayed, sorted by their priority (see 3.3).[13] Each topic has a status, which shows whether the topic has been *passed*, is *in-progress*, or *forthcoming*. For each *in-progress* topic, one educational resource is recommended (displayed). Besides accessing (and learning) the content, the learner has the following options with respect to the recommended learning content:

- **Change**: If the learners are not satisfied with the content for some reason (e.g., it is not relevant, instruction does not fit the preference, the format of the content is not preferred, low technical quality of the video/audio/text), they can replace the presented learning content, with another one addressing the same topic, at the same level. Thus, the recommendation engine records this *Change command* as an instance of feedback with a minimal value. At the same time, it updates the learners' short-term preference vector as described in Section 3.5.1, and provides an alternative educational resource, on the basis of the updated vector.
- **Done**: When a learner completes a specific learning resource, they can indicate that with the *Done* button, and optionally rate the learning content on a 5-point rating scale. The learner's profile is automatically updated based on this rating, as described in Section 3.5.1. Learners can also indicate whether they would like another learning content on the same topic, or whether they would like to try to progress to the next learning topic(s) related to a particular skill target by (successfully) taking a progress assessment (see 3.4.4).

[8] As an example, for a *Short* content, we set the Length-Short feature to 1, and the Length-Medium and Length-Long features to 0.

[9] https://github.com/ali-faraji90/edoer/blob/main/Files/Demo.mp4?raw=true.

[10] https://github.com/ali-faraji90/edoer/blob/main/Files/RegistrationForm.png.

[11] https://github.com/ali-faraji90/edoer/blob/main/Files/GoalSelection.png.

[12] https://github.com/ali-faraji90/edoer/blob/main/Files/Preferences.png.

[13] https://github.com/ali-faraji90/edoer/blob/main/Files/Curriculum.png.

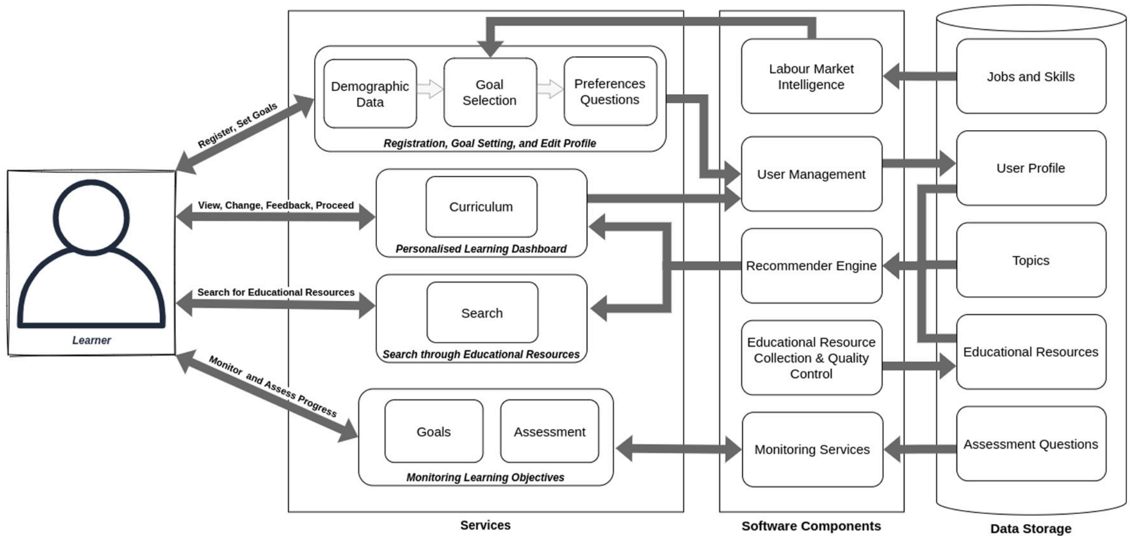

Fig. 1. Interaction between different parts of our prototype dashboard to provide the required services.

3.6.3. Search page

In order to address *Service_1* in Section 3.1.4, we provided a straightforward and simple way for learners to search through all of the open and available learning resources that are accessible on our platform. Search results (a particular learning resource) can be added to the curriculum page, and they are displayed as extracurricular learning resources.

3.6.4. Monitoring learning objectives

To address *Service_4* (see 3.1.4), we implemented a Goal Page for learners to gauge their learning progress towards their skills targets. The page, therefore, provides detailed information on the number of completed learning topics for each skill.[14] It should be mentioned that changing the target job, or removing a skill from the skill targets may remove incomplete skill training curricula from the curriculum page. However, learners can view both the new or updated target skills, and also those skills that have been removed or are incomplete. Learners can also reactivate incomplete skills by means of a simple click.

Moreover, monitoring learning processes require attention to *Assessment*, as we discussed under *Service_4* in Section 3.1.4. For this reason, we deployed an Assessment Page to structure and keep track of all skill assessments explained in Section 3.4.4. On this page, learners can see and engage with comprehensive skill assessments for each of their target skills. Furthermore, to plot individual performance on skill assessments over time, a skill assessment history for each target skill is provided as a graph.

3.6.5. History page

This page contains all learning resources that have previously been recommended to the learner. This gives learners the opportunity to review any of these learning resources at will. Content on the history page is also categorized based on learners' target skills and topics. Learners can also find information about their feedback regarding learning content, including a timestamp of completion.

3.6.6. Profile page

This page provides access (read and edit) to all the data we collected during the registration process (see Section 3.6.1) and beyond. This includes all demographic data, the target job, target skills, and learning preferences.

[14] https://github.com/ali-faraji90/edoer/blob/main/Files/Goals.png.

4. Validation

In an effort to demonstrate the effectiveness of our proposed online open educational resource recommender (*eDoer*), below we report on a randomized experiment carried out with the explicit aim of having real users interact with our prototype. The experiment was conducted to support the internal validity of our system, by answering the question of whether engagement with our system results in improved knowledge acquisition. In this part, we showcase the methods and results of this validation step.

4.1. Objective

As mentioned earlier, to more formally evaluate *eDoer* with a particular focus on evidencing the internal validity of our inferences pertaining to the effectiveness of *eDoer* in imparting knowledge in a sample of students, we set out to conduct an experiment in the context of learning about statistics. Specifically, we formulated and tested the following hypotheses, which were, by and large, premised on the fact that we specifically developed eDoer to address the most important requirements signaled by key stakeholders (see 3.1). Relying on an experimental design, enhanced our ability to rule out alternative explanations for any observed effects.

1. ***Hypothesis 1***. Using *eDoer*, as opposed to self-directed online search for open educational resources to learn about basic statistics, has a positive effect on knowledge of basic statistics.
2. ***Hypothesis 2***. Having *eDoer* provide personalized recommendations in terms of educational format (webpage, video, book, slide), length (short, medium, long), level of detail (low detail, medium detail, high detail), and content type (including example/theory or not), as opposed to having *eDoer* provide random content (from the quality controlled materials), has a positive effect on knowledge of basic statistics.

Although these hypotheses are limited in their breadth and coverage of the eDoer system, we feel they address the core functionalities/requirements that we wanted to evidence at this stage.

4.2. Procedure

For this experiment, we used the *Prolific* platform[15] which is a commercial service provider for connecting researchers with participants. In light of financial constraints associated with compensating respondents for their time (we paid each respondent 15.76 British pounds – approximately 21.74 US dollars – for their time and effort), we set out to collect high-quality learning data from a total of 150 participants. For this purpose, we decided to recruit a total number of 175 users as we predicted that we might need to remove some of the participants' data for different reasons (such as technical problems and/or missing data).

We selected "*Basic statistics for engineers*" as the target skill for this study and ran our topic extraction method on it which resulted in the following seven topics: *1. central tendency measures* (i.e., *mean, median, mode*), *2. variance and standard deviation, 3. covariance and correlation, 4. conditional probability and independent variables, 5. normal distribution, 6. linear regression*, and 7. *hypothesis testing, p-value, and confidence interval*. The reason that we selected this particular skill was to target a fundamental (engineering-related) skill while at the same time ensuring the availability of open educational resources for those people assigned to the control group (who would not be engaging with eDoer).

In order to take part in this study, the potential participants needed to complete the following steps:

1. **Step 1: pretest**. In the first step, all users participated in a pretest[16] on "*Basic statistics for engineers*" that assessed prior knowledge of the aforementioned seven topics. The test included seven questions (one question per topic) which were selected through a discussion between three experts. The experts were also asked to define the required time for each question in a way that if a participant knew a topic, he or she would have enough time to answer the question in the allotted time period.
   After completing the pretest, participants were randomly assigned to one of the following groups (to which they remained blind):
   - **Group 1**: Self-directed learning using online searches, but without any support from *eDoer*
   - **Group 2**: Learning through *eDoer* without personalized recommendations
   - **Group 3**: Learning through *eDoer* with personalized recommendations
2. **Step 2: Learning process**. In this step, the participants were granted 105 min (15 min per topic) and instructions (according to their assigned group) to study the aforementioned topics in order to be able to answer a new set of questions. The questions were on the same topics as the pretest and within the same level of difficulty. The instructions were as follows:
   - **Group 1**: In the learning process, the participants were presented with the 7 extracted topics for a finer grain searchability. They were free to engage with any type of educational content they could find (e.g., through online searches, reading books, and watching educational videos).
   - **Group 2 & 3**: These groups received simple instructions on 1. how to log in to *eDoer* using information from pre-registered new test-users, 2. fill the preference form on *eDoer*,[17] and 3. adding the skill "*Basic statistics for engineers*" to their learning profile. Subsequently, they were directed to the curriculum page to start studying each of the topics for the target skill within the defined time period.
3. **Step 3: posttest**. After the learning process, all groups were directed the posttest which included the same number of questions, on the same topics, and with the same level of difficulty.[18] This set of questions was also differently timed in the same manner that we did for the pretest.
4. **Step 4: Feedback survey**. Finally, all participants completed a short survey to provide us with feedback. *Group 1* received a survey.[19] about the steps they took to learn the topics on their own. *Groups 2* & *3* received a survey[20] about their experiences using *eDoer* Also, all groups were asked a question about their impression of the study in general.

Upon examining the data, we decided to remove 14 participants from our study as they had 1 (or less than 1) correct answers from all 14 questions. We did this to exclude those respondents who were not participating seriously in our experiment. Also, we removed the data of 5 participants data because of technical issues they had faced during the study. In the end, *Group 1* consisted of *53* participants, *Group 2* of *50* participants, and *Group 3* of *53* participants.

4.3. Measures

We calculated both scores (i.e., pretest and posttest scores) for each individual participant as the number of correct answers divided by the total number of questions per test. Subsequently, we computed our first measure ***progress score*** by subtracting, for each participant, the pretest result from the posttest result.

Additionally, through *Step 4* (*Feedback survey*), we collected participants' opinions on a 5-point scale (1: lowest to 5: highest) on the following items and converted their ratings into a number between *0.0* and *1.0* (i.e., 1 as 0.0, 2 as 0.25, 3 as 0.5, 4 as 0.75, 5 as 1.0):

- *Group 1*:
  - Availability of educational content
  - Quality of educational content
  - Satisfaction with the Prolific experiment
- *Groups 2* & *3*:
  - Personalization of content
  - Quality of educational content
  - Satisfaction with the Prolific experiment
  - Suggesting *eDoer* to other learners

Finally, to quantify learner's overall satisfaction with *eDoer*'s recommendations, we decided to collect the *Evaluative Ratings* (on a 5 point-scale) for the recommended educational materials.

4.4. Analytical procedures and results

On the pretest,*Group 1*, *Group 2*, and *Group 3* achieved average scores of *0.22*, *0.24*, and *0.20*, respectively. As expected, the pretest showed that most participants had no previous experience with *Statistics* before the experiment as their scores appear to reflect random responding. Also on the posttest, *Group 1*, *Group 2*, and *Group 3* achieved an average score of *0.34*, *0.42*, and *0.42*, respectively. Based on the pretest and posttest scores, we calculated our first measure as *progress score* which showed how each group improved their knowledge

[15] https://www.prolific.co.

[16] https://uvafeb.eu.qualtrics.com/jfe/form/SV_5AudD6pyhqWb5vU.

[17] Although Group 2 were not receiving personalized material, they also filled out the preference form as they had not any information about which group they were assigned to.

[18] https://uvafeb.eu.qualtrics.com/jfe/form/SV_4Sl8QGDg5AtECSq.

[19] https://docs.google.com/forms/d/e/1FAIpQLSeEV5ekM6rAn_s0AscxTawgbVPm3eXjhwfF3Vjrqos_2HmnUg/viewform?usp=sf_link.

[20] https://tib.eu/umfragen/index.php/887411?lang=en.

Table 5
Results of the eDoer evaluation experiment.

Measures	Mean (out of 1)			Standard deviation		
	Group 1	*Group 2*	*Group 3*	*Group 1*	*Group 2*	*Group 3*
Progress-score	0.12	0.18	**0.22**	0.20	0.19	0.17
Availability of educational content	0.56	–	–	0.16	–	–
Quality of educational content	0.64	0.75	**0.82**	0.27	0.22	0.20
eDoer recommendations' rating	–	0.79	**0.87**	–	0.17	0.16
Satisfaction from the experiment	0.74	**0.90**	**0.90**	0.23	0.14	0.14
Suggesting *eDoer* to other learners	–	0.74	**0.76**	–	0.23	0.22

in the target skill on average. This measure was ***0.12*** for ***Group 1***, ***0.18*** for ***Group 2***, and ***0.22*** for ***Group 3***. As one can see *Group 3*, which benefited from both *eDoer* and personalized recommendations, showed the most improvement. *Group 2* which benefited from *eDoer* but received random (non-personalized) recommendations also showed some degree of improvement. Finally, and as expected, *Group 1* which did not engage with *eDoer* had the lowest *progress score*.

To formally test our hypotheses, a one-way ANCOVA[21] was conducted. After controlling for the pretest scores, there was a statistically significant difference in posttest scores between the groups of learners, $F(1,152) = 11.202, p < 0.001$. Further investigation through pairwise comparison of estimated means showed that there was a statistically significant difference $t(152) = 2.31, p < .05$ between the posttest scores of the group receiving eDoers' non-personalized recommendation ($M = 2.91, SD = 1.18$) and the group of self-directed learners ($M = 2.38, SD = 1.16$). Furthermore, our findings also show a significant difference $t(152) = 2.49, p < .05$ in test scores between self-directed learners and learners receiving eDoers' personalized recommendations ($M = 2.98, SD = 1.27$). However, there was no significant difference between the posttest scores of the groups receiving non-personalized or personalized recommendations from eDoer $t(152) = 0.137, p = .892$.

In support of hypothesis 1, our findings show that participants who used eDoer without personalization attained significantly higher scores on the posttest than participants who engaged in self-directed learning (i.e., those who did not use eDoer). Unsurprisingly, and further supporting hypothesis 1, participants who used eDoer with personalization also attained higher scores on the posttest than participants who engaged in self-directed learning. In contrast, no support was found for Hypothesis 2, in that there appeared to be no significant difference in the posttest scores between those receiving personalized recommendations and those receiving non-personalized recommendations, again after controlling for scores on the pretest. To account for capitalization on chance, we reran the pairwise comparisons of estimated means applying a more conservative Bonferroni correction. The results of these analyses indicated a significant difference in the expected direction between self-directed learners and learners receiving personalized recommendations $t(152) = 2.49, p < .05$ but the difference between self-directed learners and learners receiving non-personalized recommendations failed to reach statistical significance $t(152) = 2.31, p = .066$, even though it was in the expected direction. It should be noted, however, that the Bonferroni correction has been criticized for being overly strict.

Table 5 shows the results of the other measures incorporated in our study for each group. In eyeballing these data, it is noteworthy that ratings provided are most favorable for the personalized version of eDoer, followed by the non-personalized version, and finally the self-directed learning group. Moreover, the fact that 75% of the participants are willing to recommend *eDoer* to other learners, reflects their positive attitudes towards the *eDoer* platform.

[21] We also used the *Bayesian* analysis [46] to test both of our hypotheses. The reason that we also ran *Bayesian* hypothesis testing was to serve the interests of those who purport that Bayesian methods are superior [46]. However, the results did not change the conclusions we derived based on the traditional t-test.

5. Discussion

5.1. Summary of the findings

Our system aims to support learners through labor market driven intelligent models that (1) match jobs with their required skills 3.2, (2) decompose skills into meaningful components (topics) 3.3, and (3) recommend high-quality open educational content to cover each topic 3.4.2, as the key required features for learners based on the outcome of our requirement analysis (See *RQ1* and *RQ2*). Moreover, we showed that by using our 1. topic based, 2. metadata based, and 3. quality based prediction models, we can filter out the low quality educational materials in order to recommend quality resources to the learners (See *RQ3*). Using the aforementioned components, we implemented our recommender prototype and made it accessible for learners through a dashboard (See *RQ4*). We also evaluated our prototype through an experiment in the context of a fundamental engineering skill (i.e., *Basic Statistics*). This validation indicated that learners benefited from receiving recommendations (see Hypothesis 1), and particularly so when such recommendations were personalized as evidenced by higher scores on the posttest of *group 2* and *3* combined (eDoer groups), as compared with the self-directed learning *group 1*.

The hypothesized (see hypothesis 2) findings for the difference between learners who received personalized content as opposed to those who received non-personalized content (i.e., randomly selected content) were less convincing, in that our most conservative test of this hypothesis, failed to reach statistical significance. Having said that, we should remind ourselves that personalization is but a feature of our tool, and that based on the findings for hypothesis 1, we may conclude that it made a difference to students' learning, despite the effect pertaining to the difference between the personalized and non-personalized group not reaching statistical significance. When it comes to the lack of support for hypothesis 2, one explanation is that both the personalization and the non-personalization group received quality content, and that in some instances members of the non-personalization group may in fact have received personalized content by chance (according to the limited number of educational resources that were offered for each topic). This would mean that those members contaminated what ought to have been an all non-personalization group with some degree of personalization, therewith reducing the effect size. Note that this explanation does not work for hypothesis 1 because we are certain that none of the members in the control group can have made use of our tool (*eDoer*), hence preventing such undesirable diffusion of treatment.

5.2. Limitations and future work

The initial results of our validation are promising in that they seem to indicate that engagement with eDoer, particularly when it offers personalized recommendations pertaining to statistics, appears to contribute to knowledge acquisition. Nevertheless, and as with all research, clearly there are a number of limitations that need to be acknowledged. First off, the sample size of our requirements gathering was quite limited, in that learners in different contexts, at different levels, and of different ages, and from different cultures may have different requirements that we have yet to learn about. Furthermore, people with (learning) disabilities also have needs that are not addressed by the current rendition of the system. A related challenge we faced in the requirements gathering process was how to reconcile free text input (in which we could qualitatively identify all the different requirements that learners felt needed to be addressed) with the ranking of these same requirements (with which we could determine which requirements were most important). Future work must be carried out to identify and address these needs, particularly if eDoer is to contribute to meeting the United Nations Sustainable Development Goal of providing inclusive and equitable quality education and promote lifelong learning opportunities for all, as was suggested in the introduction.

Despite the positive validation results there are also several issues that are noteworthy with regard to our experiment. Our validation comprised a limited sample of learners, studying but a single topic for a very limited amount of time. It remains to be seen whether results will be equally promising when eDoer is deployed in different contexts (for instance with unpaid learners, refugees, and/or those seeking to qualify themselves for a new occupation), in other cultures, with other learning content, and for a longer duration. To illustrate our point about duration, when we examined well-known courses on basic statistics from *Standford University*,[22] *The University of Amsterdam*,[23] and *Khan Academy*,[24] for instance, we determined that there students spend an average of 10 h (600 min) to master the aforementioned topics on basic statistics. Given that the current study established a treatment effect for what constituted but a very limited 'dosage' of training, strengthens us in the belief that stronger effects can be booked with trainings of greater duration and depth. Clearly, however, future research is needed to further develop and evidence this tool, with different samples, different topics, and training of greater durations.

In addition to training duration, one may also wonder about the longer term retention of that which was learned, in that our posttest was administered quite soon after the training. Future research will need to examine the extent to which that which was learnt is retained over time. Here too, however, we feel that retention is only likely to improve with trainings of greater duration.

Based on the feedback and the lessons we learned during the prototype development process, we also conclude that more work needs to be done on the personalization and scalability components of our prototype. Specifically, to personalize the learning experience, we collected several initial personal features from learners (i.e. length, level of detail, learning strategy, and content format 3.5.1). However, this still needs to be extended to describe the learners' context in a fine-grained manner. Therefore we see value in capturing more preference features in the future, such as language preferences, preferred authors, location, or sensory information on learners' cognitive and mental state (e.g., tiredness, well-being).

Moreover, currently, we use long-term and short-term vectors to plot learner preferences. At the moment, it puts more emphasis on their recent feedback about learning content they studied 3.5.1. In the user profile, however, learners can edit their long-term vector (the basis for recommendations (see 3.5.2)) directly, which overwrites their preference score, computed by our model, based on actual learner feedback and behavior (see 3.5.1). Therefore, we will need to fine-tune this scoring algorithm by, for instance, providing an option for learners to decide about the balance between their long-term and short-term vectors.

To arrive at a scalable open educational recommender system, we need to address two further issues: (1) Intelligent models, which we use in our educational content matching and content quality prediction steps, need to be very precise and accurate. Our algorithms sometimes cannot provide the level of accuracy, which is needed for automatic decisions on content inclusion. Therefore, currently, we need to handle the errors that are produced by the models and minimize their impact manually. (2) Extracting properties from educational resources (currently done by manual labeling) is a time-consuming, and error-prone activity. To tackle these problems, we plan to move towards a crowdsourcing based quality monitoring and labeling strategy. This will not only check the output of our intelligent models but will also improve our models (or help in building new models) based on the participants' (crowd) opinion.

[22] https://www.coursera.org/learn/stanford-statistics.

[23] https://www.coursera.org/learn/basic-statistics.

[24] https://www.youtube.com/watch?v=uhxtUt_-GyM.&list=PL1328115D3D8A2566&ab_channel=KhanAcademy.

6. Conclusion

To remain employable, learners continuously need to master skills and topics that are relevant for their desired jobs in a dynamically changing labor market. We initiated the work reported in this manuscript by conducting a requirement analysis to extract the learners' need for such a learning environment. Based on the results of our analysis, we designed and implemented a system, called *eDoer*, that helps learners to set their learning goals and to receive a personalized learning path towards their goals. These learning paths contain high-quality educational materials which have passed through our automatic quality control models (i.e. topic based, metadata based, and quality based prediction models). We evaluated our prototype system through an experiment in the context of a fundamental engineering skill (i.e. *Basic Statistics*). This validation showed tentative support for our first hypothesis, indicating that learners who used our system, performed better on a posttest than those learners engaging in self-directed learning. The findings for the learners who received non-personalized (i.e., randomly selected content) were less convincing, in that our most conservative test of this hypothesis, which was about the difference between the personalized and non-personalized group, failed to reach statistical significance.

CRediT authorship contribution statement

Mohammadreza Tavakoli: Methodology, Software, Validation, Data Curation, Writing – Original Draft. **Abdolali Faraji:** Methodology, Software, Data Curation, Writing - Original Draft, Visualization. **Jarno Vrolijk:** Methodology, Validation, Writing - Original Draft. **Mohammadreza Molavi:** Validation. **Stefan T. Mol:** Conceptualization, Validation, Supervision, Validation, Writing - Review & Editing. **Gábor Kismihók:** Conceptualization, Supervision, Writing - Review & Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge the financial support from the following projects that have helped in developing the eDoer platform: ADSEE - Applied Data Science Educational Ecosystem, European Commission - Erasmus Plus Programme, 2019-1-HR01-KA203-060984; OSCAR - Online, open learning recommendations and mentoring towards Sustainable research CAReers, European Commission - Erasmus Plus Programme, 2020-1-DE01-KA203-005713; BIPER - Business Informatics Programme Reengineering, European Commission - Erasmus Plus Programme, 2020-1-HU01-KA226-HE-093987; ADAPT - Implementation of an Adaptive Continuing Education Support System in the Professional Field of Nursing German Federal Ministry of Education and Research BMBF - INVITE 21INVI0501; WBsmart - AI-based digital continuing education space for elderly care, German Federal Ministry of Education and Research BMBF - INVITE 21INVI2101.

References

[1] United Nations, Goal 4 | department of economic and social affairs, 2020, https://sdgs.un.org/goals/goal4.
[2] N. Sclater, Open educational resources: Motivations, logistics and sustainability, in: Content Management For E-Learning, Springer, 2011, pp. 179–193.
[3] V.B. Kobayashi, S.T. Mol, H.A. Berkers, G. Kismihók, D.N. Den Hartog, Text mining in organizational research, Organ. Res. Methods 21 (3) (2018) 733–765.
[4] M. Tavakoli, G. Kismihok, S.T. Mol, Labour market information driven, personalized, OER recommendation system for lifelong learners, in: International Conference On Computer Supported Education (CSEDU), SciTePress, 2020.
[5] F. Wang, Z. Jiang, X. Li, G. Li, Cognitive factors of the transfer of empirical engineering knowledge: A behavioral and fNIRS study, Adv. Eng. Inform. 47 (2021) 101207.
[6] E. Colombo, F. Mercorio, M. Mezzanzanica, Applying machine learning tools on web vacancies for labour market and skill analysis, in: Terminator Or The Jetsons? The Economics And Policy Implications Of Artificial Intelligence, 2018.
[7] V. Castello, E. Flores, M. Gabor, J. Guerrero, M. Guspini, J. Luna, L. Mahajan, K. McGartland, I. Szabo, F. Ramos, Promoting dynamic skills matching: challenges and evidences from the smart project, in: INTED2014 Proceedings, Citeseer, 2014, pp. 2430–2438.
[8] X. Li, Z. Jiang, Y. Guan, G. Li, F. Wang, Fostering the transfer of empirical engineering knowledge under technological paradigm shift: An experimental study in conceptual design, Adv. Eng. Inform. 41 (2019) 100927.
[9] I. Wowczko, Skills and vacancy analysis with data mining techniques, in: Informatics, vol. 2, (4) Multidisciplinary Digital Publishing Institute, 2015, pp. 31–49.
[10] V.B. Kobayashi, S. Mol, G. Kismihok, Labour market driven learning analytics, J. Learn. Anal. 1 (3) (2014) 207–210.
[11] M.M. McGill, Defining the expectation gap: a comparison of industry needs and existing game development curriculum, in: Proceedings Of The 4th International Conference On Foundations Of Digital Games, ACM, 2009, pp. 129–136.
[12] M. Tavakoli, A. Faraji, S.T. Mol, G. Kismihók, Oer recommendations to support career development, IEEE Frontiers In Education (FIE) (2020).
[13] M.d.C. Saraiva, et al., Relationships among educational materials through the extraction of implicit topics: Relacionamentos entre materiais didáticos através da extração de tópicos implícitos. [sn], 2019.
[14] M. de Carvalho Saraiva, C.B. Medeiros, Finding out topics in educational materials using their components, in: 2017 IEEE Frontiers In Education Conference (FIE), IEEE, 2017, pp. 1–7.
[15] M. Molavi, M. Tavakoli, G. Kismihók, Extracting topics from open educational resources, in: European Conference On Technology Enhanced Learning, Springer, 2020, pp. 455–460.
[16] J. Djumalieva, C. Sleeman, An open and data-driven taxonomy of skills extracted from online job adverts, in: Developing Skills In A Changing World Of Work: Concepts, Measurement And Data Applied In Regional And Local Labour Market Monitoring Across Europe, 2018, p. 425.
[17] J. Wang, J. Xiang, K. Uchino, Topic-specific recommendation for open education resources, in: International Conference On Web-Based Learning, Springer, 2015, pp. 71–81.
[18] M. Tavakoli, M. Elias, G. Kismihok, S. Auer, Quality prediction of open educational resources - a metadata-based approach, IEEE, 2020,
[19] M. Tavakoli, S. Hakimov, R. Ewerth, G. Kismihok, A recommender system for open educational videos based on skill requirements, IEEE, 2020,
[20] G. Sun, T. Cui, D. Xu, J. Shen, S. Chen, A heuristic approach for new-item cold start problem in recommendation of micro open education resources, in: International Conference On Intelligent Tutoring Systems, Springer, 2018, pp. 212–222.
[21] J. Chicaiza, N. Piedra, J. Lopez-Vargas, E. Tovar-Caro, Recommendation of open educational resources. An approach based on linked open data, in: Global Engineering Education Conference, IEEE, 2017, pp. 1316–1321.
[22] S. Wan, Z. Niu, An e-learning recommendation approach based on the self-organization of learning resource, Knowl.-Based Syst. 160 (2018) 71–87.
[23] G. Sun, T. Cui, G. Beydoun, S. Chen, F. Dong, D. Xu, J. Shen, Towards massive data and sparse data in adaptive micro open educational resource recommendation: a study on semantic knowledge base construction and cold start problem, Sustainability 9 (6) (2017) 898.
[24] A.R. Fabbri, I. Li, P. Trairatvorakul, Y. He, W.T. Ting, R. Tung, C. Westerfield, D.R. Radev, Tutorialbank: A manually-collected corpus for prerequisite chains, survey extraction and resource recommendation, 2018, arXiv preprint arXiv:1805.04617.
[25] A. García-Floriano, A. Ferreira-Santiago, C. Yáñez-Márquez, O. Camacho-Nieto, M. Aldape-Pérez, Y. Villuendas-Rey, Social web content enhancement in a distance learning environment: intelligent metadata generation for resources, Int. Rev. Res. Open Distributed Learn. 18 (1) (2017) 161–176.
[26] S. Basu, Y. Yu, R. Zimmermann, Fuzzy clustering of lecture videos based on topic modeling, in: 2016 14th International Workshop On Content-Based Multimedia Indexing (CBMI), IEEE, 2016, pp. 1–6.
[27] H. Shukla, M. Kakkar, Keyword extraction from educational video transcripts using nlp techniques, in: 2016 6th International Conference-Cloud System And Big Data Engineering (Confluence), IEEE, 2016, pp. 105–108.
[28] K. Horita, F. Kimura, A. Maeda, Automatic keyword extraction for wikification of east Asian language documents, Int. J. Comput. Theory Eng. 8 (1) (2016) 32.
[29] S. Aryal, A.S. Porawagama, M.G.S. Hasith, S.C. Thoradeniya, N. Kodagoda, K. Suriyawansa, Using pre-trained models as feature extractor to classify video styles used in MOOC videos, in: 2018 IEEE International Conference On Information And Automation For Sustainability (ICIAfS), IEEE, 2018, pp. 1–5.
[30] P. Király, M. Büchler, Measuring completeness as metadata quality metric in Europeana, in: 2018 IEEE International Conference On Big Data (Big Data), IEEE, 2018, pp. 2711–2720.
[31] A.R. Pelaez, P.P. Alarcon, Metadata quality assessment metrics into OCW repositories, in: Proceedings Of The 2017 9th International Conference On Education Technology And Computers, ACM, 2017, pp. 253–257.
[32] A. Romero-Pelaez, V. Segarra-Faggioni, P.P. Alarcon, Exploring the provenance and accuracy as metadata quality metrics in assessment resources of OCW repositories, in: Proceedings Of The 10th International Conference On Education Technology And Computers, ACM, 2018, pp. 292–296.
[33] A. Romero-Pelaez, V. Segarra-Faggioni, N. Piedra, E. Tovar, A proposal of quality assessment of OER based on emergent technology, in: 2019 IEEE Global Engineering Education Conference (EDUCON), IEEE, 2019, pp. 1114–1119.
[34] X.N. Lam, T. Vu, T.D. Le, A.D. Duong, Addressing cold-start problem in recommendation systems, in: Proceedings Of The 2nd International Conference On Ubiquitous Information Management And Communication, ACM, 2008, pp. 208–211.
[35] M. Khobreh, F. Ansari, M. Fathi, R. Vas, S.T. Mol, H.A. Berkers, K. Varga, An ontology-based approach for the semantic representation of job knowledge, IEEE Trans. Emerg. Top. Comput. 4 (3) (2015) 462–473.
[36] D. Smith, A. Ali, Analyzing computer programming job trend using web data mining, Issues Inf. Sci. Inf. Technol. 11 (1) (2014) 203–214.
[37] P.G. Lovaglio, M. Cesarini, F. Mercorio, M. Mezzanzanica, Skills in demand for ICT and statistical occupations: Evidence from web-based job vacancies, Stat. Anal. Data Min. ASA Data Sci. J. 11 (2) (2018) 78–91.
[38] E.M. Sibarani, S. Scerri, C. Morales, S. Auer, D. Collarana, Ontology-guided job market demand analysis: a cross-sectional study for the data science field, in: Proceedings Of The 13th International Conference On Semantic Systems, ACM, 2017, pp. 25–32.
[39] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, L. Zhao, Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey, Multimedia Tools Appl. 78 (11) (2019) 15169–15211.
[40] M. Tavakoli, M. Elias, G. Kismihók, S. Auer, Metadata Analysis of Open Educational Resources, in: LAK21: 11th International Learning Analytics And Knowledge Conference, 2021, pp. 626–631.
[41] T.F. Hawk, A.J. Shah, Using learning style instruments to enhance student learning, Decis. Sci. J. Innov. Educ. 5 (1) (2007) 1–19.
[42] Wikipedia, Dot product, 2020, https://en.wikipedia.org/wiki/Dot_product.
[43] Mozilla, Responsive design, 2020, https://developer.mozilla.org/en-US/docs/Learn/CSS/CSS_layout/Responsive_Design#responsive_design.
[44] Google, Material design, 2020, https://material.io/.
[45] Introjs, Introduce users to your product, 2020, https://introjs.com/.
[46] J.K. Kruschke, BayesIan estimation supersedes the t test., J. Exp. Psychol. General 142 (2) (2013) 573.