Connecting Learning Material and the Demand of the Job Market Using Artificial Intelligence

Darragh Carroll and Tim Schlippe

IU International University of Applied Sciences, Germany.

Email: tim.schlippe@iu.org;

Abstract

To optimally prepare students for jobs, it is often useful to match the content of the learning material with the needs of the current job market. On the other hand, it can motivate students and give them inspiration for their future careers to see what exciting jobs they can acquire if they learn the learning material. For these reasons, we have explored artificial intelligence methods to compare the learning content with the skills required in current job postings and provide feedback to both the teachers and the students in the form of reports. Our best system combines Sentence-BERT [1], K-Means [2], DBSCAN [3], TF-IDF [4] and SVMs [5] to build a predictive model capable of identifying job market skills within learning content with an accuracy of 94.2% and and F-score 86.9%. The results of our subsequent survey demonstrate that our reports for students are understandable, show the learning content's purpose, motivate, accelerate the learning process, as well as give them job market and learning path information. Additionally, our report for teachers is understandable, shows importance of present and missing skills in the learning content, accelerates and specifies the learning content creation, and helps them advise students.

Keywords: Natural Language Processing, Artificial Intelligence, AI in Education, Recommender System

1 Introduction

Education plays a key role in economic prosperity, societal well-being, and technological innovation. It is an important aspect of today's world and can help humanity to solve many problems—both present and future. Adult learning and education is now treated using a human rights-based approach to align with goal 4 of the United

Nations' Sustainable Development Agenda [6]. It is important to continuously equip the population with the necessary skills to succeed in the job market and drive humanity's progress toward sustainability.

The needs of the job market are in constant flux due to declining and expanding sectors [7]. New technologies and innovations are constantly changing the importance of various roles in the job market, both in a transformative and destructive manner [8]. Recent advancements have introduced completely new roles into the job market, such as the AI ethicist [9]. Other roles have had to adapt to keep pace with new technologies, such as the requirement for physicians to understand and use AI as they practice medicine [10]. Other roles, such as manufacturing, are at threat of being replaced by sophisticated automation systems [11].

The dynamic nature of the job market makes it difficult for education institutions to keep their learning curricula up-to-date. For example, many university courses are requiring a framework restructure in order to incorporate AI into their teaching strategies [12]. As technology advances, and the job market adapts, it is necessary to align the learning curricula in order for graduates to acquire relevant skills. This is important since companies are placing a greater emphasis on a candidate's skills rather than their degree [13].

In addition to maintaining relevant and up-to-date learning content, it is important to motivate students as the motivation to learn can significantly affect the learning outcome [14]. Students can become far more motivated when they understand how the learning content is relevant to them and their career goals [15]. Additionally, the self-determination theory can be exploited to enhance a student's motivation [16]: This theory states that motivation can benefit from the ability to make choices. As a result, there may be great potential in allowing the student to not only choose their study program but also choose and personalize their learning path.

Due to all theses reasons, our work incorporates natural language processing (NLP) and predictive modeling in the design of an artificial intelligence (AI) system to extract skills from the job market for various jobs and identifying where these skills can be learned in learning content. Our system's ability to highlight present/missing skills helps teachers to keep their learning content relevant and up-to-date. Our system also helps motivate students by explaining the relevancy of various sections of learning content, i.e. what skills they can learn and what jobs they can get with these skills. Additionally, our system supports students to acquire skills relevant to their desired job by recommending different sections of content where these skills can be learned.

In the next section, we will present related approaches to connect education and job market with the help of AI as well as techniques for vectorizing, clustering, searching and comparing text data. In Section 3 we will demonstrate the pipeline of our system to connect learning material and job postings. Section 4 will describe the experimental setup for finding the optimal technical implementations for our pipeline. The choices for implementation are outlined in Section 5. In Section 6 we will present our detailed analysis of the feedback from students and teachers on the reports generated with our system. We will conclude our work in Section 7 and suggest further steps.

2 Related Work

In this section we will first describe related work to connect education to the job market. Then we will look at approaches to summarize, extract and compare skills.

2.1 Connecting Education and the Job Market

Various strategies have been implemented to better connect the education industry with the job market. For example, [17] have adopted a 'link and match' approach where employers are consulted regularly to keep the learning content updated. However, employers need to be included to accurately understand the needs of the job market. Consequently, [18] combine industry survey results with emerging technology trends to better align the learning content with job market needs.

Another approach to align these two industries is to implement a flexible, modular learning experience to enable students to adapt their learning in response to changes in the job market [19]. Offering students the chance to select different courses not only enhances their enjoyment and motivation, but it can also ensure that the students graduate with the necessary skills tailored to their desired career path.

[20] propose to treat students, educational institutions and employers as stakeholders in an 'education supply chain' where their interactions are crucial to ensure the continuous flow of highly skilled graduates. Such interactions could include internships, guest lecturers, joint research projects, and advisory boards consisting of all three stakeholders. Although these person-to-person interactions can be effective in connecting education with the job market, [20] emphasize that the utilization of AI could further optimize the 'education supply chain'.

Consequently, in [21–23] we proposed an NLP pipeline to connect students, educational institutions and employers which recommends study programs to help job candidates acquire relevant skills. In this paper, however, we present an AI system to suggest parts of learning content to help students acquire relevant skills and teachers align their learning content to the job market.

2.2 Analyzing Skills with Natural Language Processing

To align skills from different text sources, NLP techniques can be used to summarize text and extract skills. For example, [24] extract and represent relevant skills by comparing knowledge graphs for job descriptions and candidate profiles. [25] use Word2Vec [26] to extract skills from CVs and automatically find the best candidate. [27] combine TF-IDF (term frequency—inverse document frequency) and vectorization to obtain a list of the best job candidates, ranked using the extracted skills. [28] use K-Means clustering to assess a job applicant's skill set and recommend the job most suited to the applicant. [29] combine K-Means and hierarchical clustering to group students into clusters depending on present/missing skills, enabling cluster-specific teaching methods. [30] implement DBSCAN to extract skills from job descriptions, even when the same skills are worded differently.

In [21–23] we introduced *SkillScanner*, which uses a combination of Sentence-BERT, DBSCAN, K-Means and vectorization to extract and represent skills from job postings, CVs and learning curricula. In this paper, we incorporate TF-IDF and SVM into the *SkillScanner* pipeline to locate skills in the titles, sentences and paragraphs of the learning content.

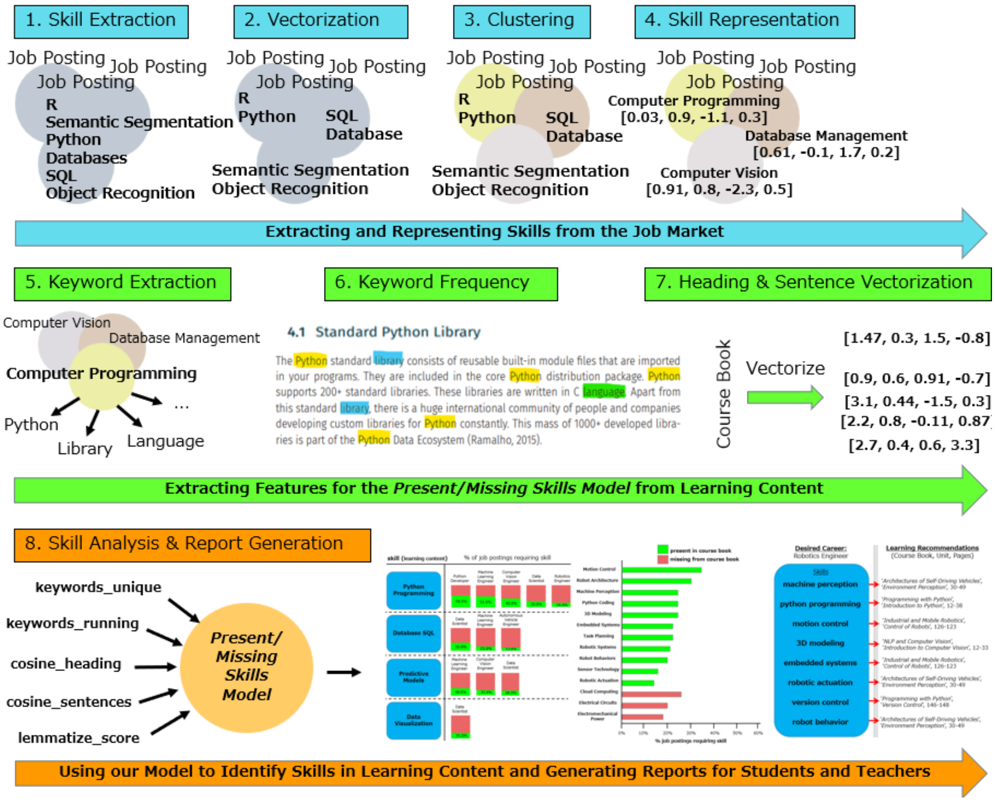

Fig. 1: Pipeline to Connect Learning Material and Job Postings.

3 Pipeline to Connect Learning Material and Job Postings

As demonstrated in Figure 1, our pipeline consists of 8 steps, which can be divided into 3 phases with respect to the data processed in the steps:

1. *Extracting and Representing Skills from the Job Market*: Skills from job postings are extracted and represented as sentence embeddings.
2. *Creating Features for the Present/Missing Skills Model from Learning Content*: The skills are searched for in the learning content.
3. *Using our Model to Identify Skills in Learning Content and Generating Reports for Students and Teachers*: A predictive model takes features which are produced

based on the outputs of the first and the second phase and predicts the job market skills which are present and missing in the learning content.

The individual steps to compare the skills required in job postings with learning material are as follows:

1. *Skill Extraction*: Extract skill requirements from each job posting (e.g. *R*, *Semenatic Segmentation*, *Python* from job postings of a Data Scientist).
2. *Vectorization*: Map skill requirements extracted from the job postings to a semantic vector space, where skills with similar meanings are closer together and skills with different meanings are farther apart.
3. *Clustering*: Cluster skill requirements to cope with the challenges of different levels of abstraction and synonyms as well as to remove outliers.
4. *Skill Representation*: After removing outliers, retrieve for each remaining skill cluster:
   (a) *skill vector*: Average vector of all word embeddings belonging to this cluster.
   (b) *skill sentence*: Sentence whose vector is closest to the *skill vector*.
   (c) *skill label*: Most frequent bigram that occurs in the skills belonging to this cluster.
5. *Keyword Extraction*: Retrieve more keywords which are specifically related to the extracted skills and may occur in the learning material.
6. *Keyword Frequency*: Compute the absolute and relative frequencies of the skill-related keywords in the learning material:
   (a) *keywords_unique*: Sum of the unique keyword types.
   (b) *keywords_running*: Sum of the running keyword tokens.
7. Map headings and each sentence in the learning content to the semantic vector space which enables to compute:
   (a) *cosine_heading*: Similarity between each vectorized heading and the *skill vectors*.
   (b) *cosine_sentences*: Similarity between each vectorized sentence and the *skill vectors*.
   (c) *lemmatize_score*: Similarity between each lemmatized sentence and the lemmatized *skill sentences* (measured as the percentage of overlapping lemmas).
8. *Skill Analysis and Report Generation*: Based on the features *keywords_unique*, *keywords_running*, *cosine_heading*, *cosine_sentences* and *lemmatize_score*, our *present/missing skills model* identifies which skills from the job postings are covered and missing in the learning content which is the basis for generating reports for students and teachers.

4 Experimental Setup

The implementation of our pipeline was carried out using Google Colab[1]. In order to find the optimal technical implementation for our pipeline to connect learning material to job postings, we used the following data for evaluation:

- Job postings: To retrieve skills from job postings for our system optimization and evaluation, we collected 100 job postings for each of the following 6 jobs

[1] https://colab.research.google.com

on Indeed.com: Autonomous Vehicle Engineer, Computer Vision Engineer, Data Scientist, Machine Learning Engineer, Python Developer, and Robotics Engineer.
- Lecture content: To obtain learning material for our system optimization and evaluation, we used 3 sections/chapters from each of the following 6 courses from our university: 'Architectures of Self-Driving Vehicles', 'NLP and Computer Vision', 'Software Engineering for Data Intensive', 'Machine Learning', 'Programming with Python', 'Industrial and Mobile Robotics'. This resulted in over 250 pages of learning material. Manual labeling was then performed to highlight present/missing skills in each section, and this labeled data was used to train and test our *present/missing skills model*.

5 Our Implementation

In this section, we will describe how we implemented the pipeline to connect learning material and job postings which was described in Section 3 with regard to the job postings and course books specified in Section 4. We share the code and the data with the research community on GitHub[2].

5.1 Extracting and Representing Skills from the Job Market

To extract, vectorize and cluster the skills listed in the job postings in step 1–3 of our pipeline, we implemented the components which gave optimal results in [21–23].

In job postings, skills are usually expressed in bullet points. Therefore, we developed keyword- and rule-based techniques to extract bullet points from job postings. To retrieve the job postings, we used the ScrapeOps free trial web scraper [31]. Thus, we retrieved 7,787 bullet points in 600 job postings to build a set of representative skills for our 6 job positions in English from Indeed.com.

Since we discovered in [21–23] that Sentence-BERT, which is based on Bidirectional Encoder Representations from Transformers (BERT) [1], gave optimal results, we used Sentence-BERT for vectorization as well.

Then we used (1) UMAP [32] to reduce the vectorized skills to two dimensions and (2) DBSCAN [3] to remove outliers in the 2-dimensional space as suggested in [21–23]. DBSCAN finds clusters in dense areas and assigns a cluster ID to all points belonging to the same cluster. Outliers—i.e., in our case, clusters that do not represent skills or are not part of a dense area—are assigned the cluster ID -1.

With those steps, we reduced the 7,787 potential skills to 6,184 skills. To find synonymous skills in the job postings, we then applied K-Means clustering with the cosine distance as the distance metric to the original 768-dimensional vectors that remained after removing outliers in the 2-dimensional space.

5.2 Extracting Features for the Present/Missing Skills Model from Learning Content

After retrieving clusters and vectors representing the skill of each cluster, we used the BeautifulSoup Python package [33] and TF-IDF to extract a set of 15 keywords for

[2] https://github.com/darragh314/SkillScanner

each skill from relevant Wikipedia pages. These keywords were then searched for and counted in 18 different sections across 6 course books. This enabled us to extract two features for our predicitve model: *keywords_unique* and *keywords_running*.

We then used Sentence-BERT [1] to vectorize the section headings and the sentences within the learning content. The cosine similarity between the vectorized section headings and the *skill vectors* from the job postings enabled us to extract the feature *cosine_heading*. Additionally, the cosine similarity between the vectorized learning content sentences and the *skill vectors* enabled us to extract the feature *cosine_sentences*.

Finally, we used lemmatization to reduce each word in the *skill sentence* to its base form. In addition, we lemmatized each sentence in the learning content and calculated the percentage of lemmatized *skill sentence* words present in the lemmatized learning content. This method has been used to successfully extract topics [34] and identify skills [35], and enabled us to retrieve *lemmatize_score* as an additional feature for our *present/missing skills model*. The five input features, *keywords_unique*, *keywords_running*, *cosine_heading*, *cosine_sentences* and *lemmatize_score* resulted in a 5D feature vector, which was input into our *present/missing skills model*.

5.3 The Present/Missing Skills Model

Once the feature extraction process was complete, we split our labeled data set in a 80:20 ratio to create the training data and test data. Of all models, SVM demonstrated the optimal predictive performance with an accuracy of 94.2% and F-score of 86.9%. Figure 2 demonstrates the feature importance of our 5 features on our 250 pages of labeled learning content in percentages. We see that *cosine_heading* (37.7%), *lemmatize_score* (24.5%), *keywords_unique* (23.6%) and *cosine_sentences* (13.8%) significantly influence the SVM model's ability to identify missing/present skills, while *keywords_running* (3.3%) does not have a big impact on the model's performance.

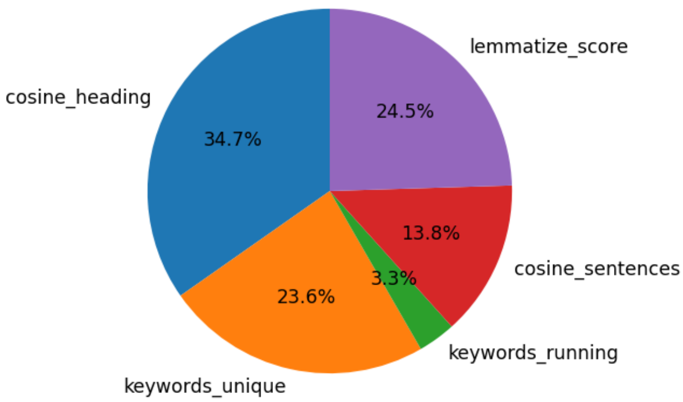

Fig. 2: Feature Importance Scores for *Present/Missing Skills Model*.

Figure 3 shows the present and missing skills in 3D space, together with the decision boundary. For the purpose of visualization, we reduced the 5D feature space to 3D using Principal Component Analysis (PCA). The false positive rate of our SVM model is 3.27%, i.e. 3.27% of missing skills were incorrectly labeled as present. The false negative rate of our SVM model is 14.52%, i.e. 14.52% of present skills were incorrectly labeled as missing. The false positives and false negatives are shown on either side of the decision boundary in Figure 3.

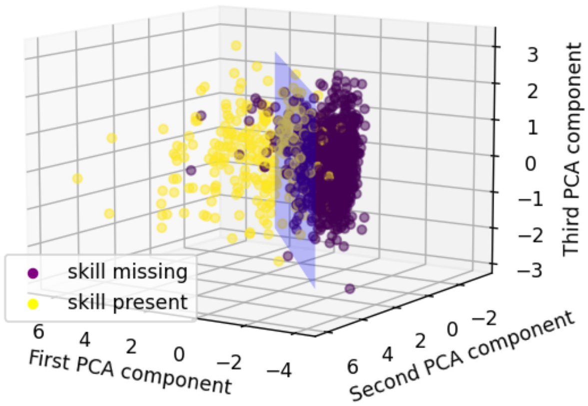

Fig. 3: SVM Decision Boundary.

5.4 Reports for Students and Teachers

Our system receives a set of skills from a job posting, compares it to the skills taught in learning content and returns the following reports.

5.4.1 LearningContent-JobMarket Report for Students

As illustrated in Figure 4, the *LearningContent-JobMarket Report for Students* shows students that they will learn skills which are important for specific jobs (e.g., Python Developer, Machine Learning Engineer, Computer Vision Engineer, Data Scientist and Robotics Engineer). The skills extracted from the learning content (e.g., unit "Python Important Libraries") are listed in blue boxes on the left (Python Programming, Database SQL, Predictive Models, etc.). Based on the skills present in the learning content, jobs which require these skills are displayed on the right. Additionally, the importance of the skills for each job is visualized by the percentage of job postings requiring that skill. The green part in each bar demonstrates how important the skill is for each job position, whereas the red part shows the percentage of job postings not requiring the skill.

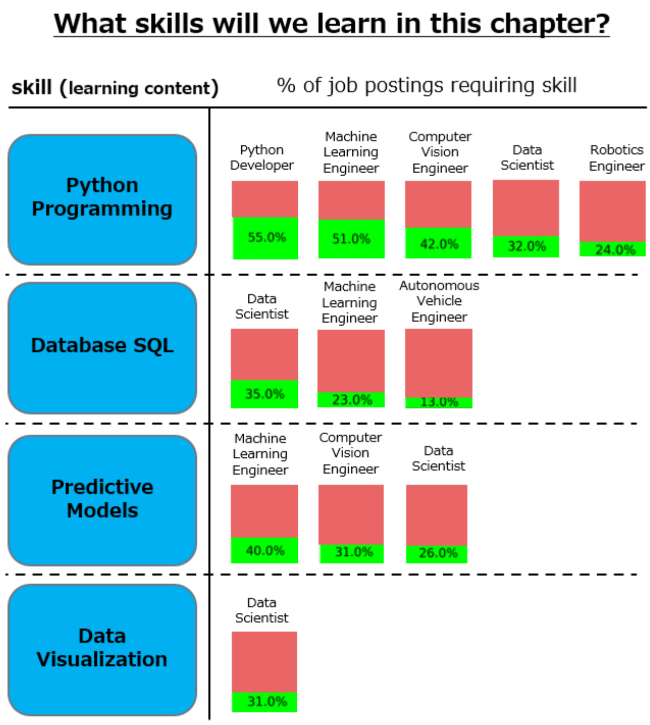

Fig. 4: *LearningContent-JobMarket Report for Students.*

5.4.2 Job-LearningPath Report for Students

As illustrated in Figure 5, the *Job-LearningPath Report for Students* recommends learning content related to a specific job (e.g., Robotics Engineer). The extracted skills are listed in the blue box on the left (machine perception, python programming, etc.). Based on the skills required in the job postings and present in the learning content, learning recommendations in terms of course books, units and pages are presented on the right.

5.4.3 LearningContent-JobMarket Report for Teachers

As illustrated in Figure 6, the *LearningContent-JobMarket Report for Teachers* shows teachers which skills, which are important for a specific job, are currently covered and missing in their learning content. Skills which are needed in the job posting and present in the learning content are displayed with a green bar. Skills that are demanded in the job posting but missing in the learning content are displayed with a red bar. The length of a bar indicates how important the skill is for each job position reflecting the percentage of job postings requiring that skill.

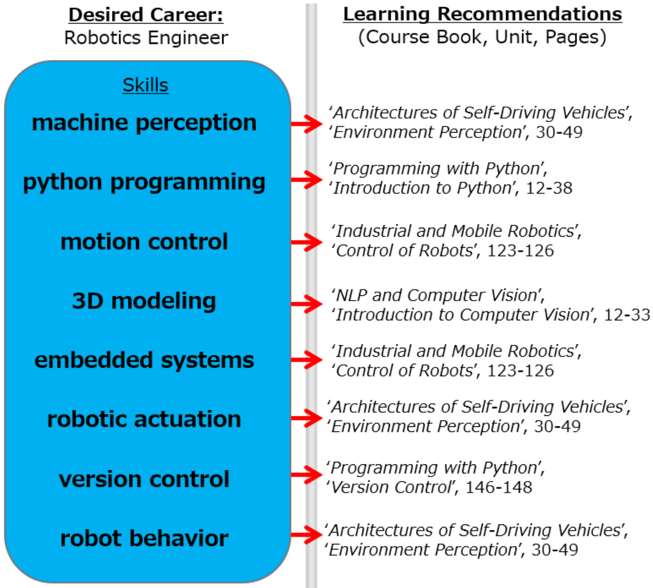

Fig. 5: *Job-LearningPath Report for Students.*

6 User Study

In this section we will describe the design and results of our survey, in which we asked for feedback on our reports.

6.1 Questionnaire Design

The goal of our study was to receive feedback on our 3 reports that can be generated with our AI system:

- *Job-LearningPath Report for Students*
- *LearningContent-JobMarket Report for Students*
- *LearningContent-JobMarket Report for Teachers*

To obtain feedback from our two target groups—students and teachers—we created a questionnaire for students where we presented the *Job-LearningPath Report for Students* and the *LearningContent-JobMarket Report for Students* and a different questionnaire for teachers where we presented the *LearningContent-JobMarket Report for Teachers* but also our two reports for students to receive pedagogical feedback. In each questionnaire, we asked questions about the reports presented.

With the questionnaire for students, our goal was investigate if with the help of these reports, students

- understand the purpose of learning a specific content (*purpose*),
- are motivated to study the learning content (*motivation*),
- can accelerate their learning process (*acceleration*),

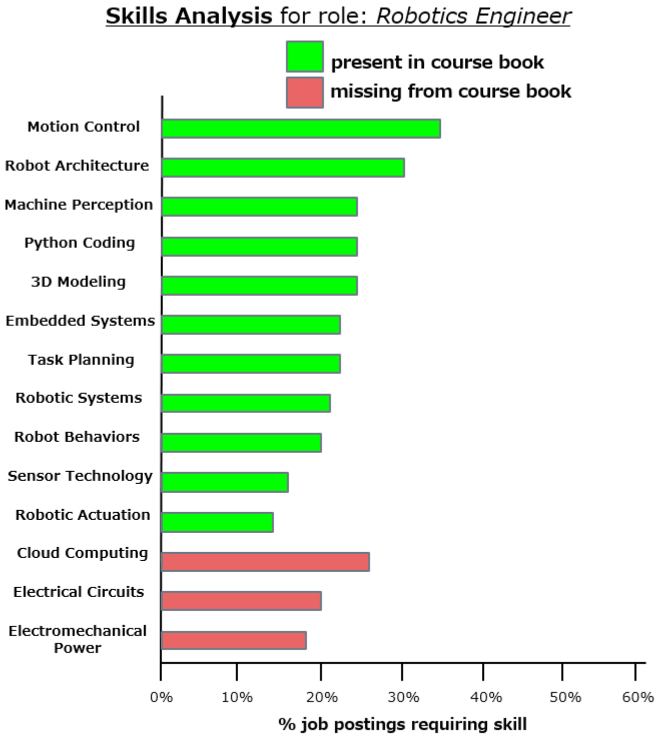

Fig. 6: *LearningContent-JobMarket Report for Teachers.*

- are informed of jobs related to the learning content (*job market information*),
- can create their own learning path to their desired job (*learning path information*),
- and find the reports' visualization understandable (*visualization*).

With the questionnaire for teachers, our goal was investigate if with the help of the *LearningContent-JobMarket Report for Teachers*, teachers

- understand the importance of the skills in the job market (*skill importance*)
- see skills from the job market that are missing in learning content (*missing skills*),
- can accelerate their study program design process (*acceleration*),
- can advise courses to students (*advise*),
- can design course books that are targeted towards specific jobs (*targeted education*),
- and find the report's visualization understandable (*visualization*).

As mentioned we also showed the teachers our two reports for students and asked them with comparable questions to analyze whether they agree with the students' feedback. Our participants evaluated most questions with a score. The score range follows the rules of a forced choice Likert scale, which ranges from (1) *strongly disagree* to (5) *strongly agree*. Each questionnaire was designed in English and German. In total, 88 participants (46 female, 42 male) filled out our questionnaires. 38 participants stated that they were students or former students (43.2%) and filled out the questionnaire for students. 50 reported working or having worked as teaching staff (56.8%) and completed the questionnaire for teachers. The students and teachers came from educational institutions located in Germany, Ireland, United Kingdom, Netherlands and the United States. They cover Science, Technology, Engineering, and Mathematics (STEM), Business and Economics, Design and Arts, Language and Linguistic, Social Sciences and other subjects.

6.1.1 Evaluation of the *LearningContent-JobMarket Report for Students*

Figure 7 and 8 demonstrate the students' and teachers' feedback on the *LearningContent-JobMarket Report for Students* with regard to the criteria *purpose*, *motivation*, *acceleration*, *job market information*, *learning path information* and *visualization*. In this report the majority of students agrees and even strongly agrees that our report is appropriate with averages of 4.61 for *purpose*, 4.37 for *motivation*, 3.89 for *acceleration*, 4.74 for *job market information*, 4.29 for *learning path information* and 4.53 for *visualization*.

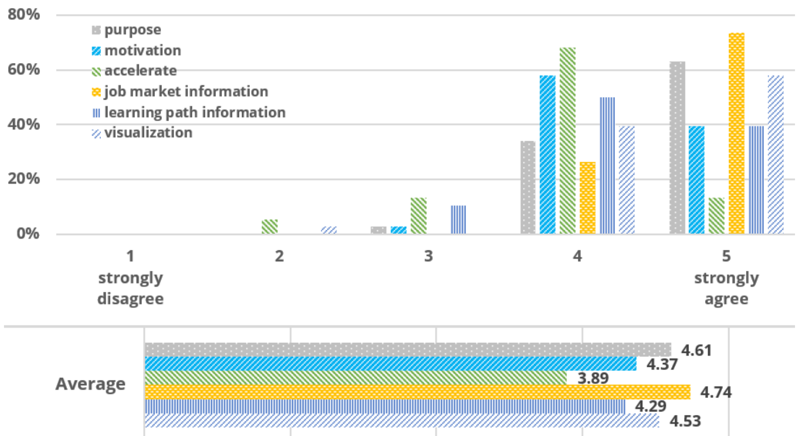

Fig. 7: Students' feedback on the *LearningContent-JobMarket Report for Students*.

Considering the teachers' feedback on the *LearningContent-JobMarket Report for Students* in Figure 8, we observe that with averages of 3.86 for *purpose*, 3.62 for *motivation*, 3.20 for *acceleration*, 4.12 for *job market information*, 3.64 for *learning path information* and 3.82 for *visualization*, the majority of teachers agrees and even strongly agrees that our report is appropriate for students regarding our evaluation criteria. While the overall average for participating students is 4.41, the overall average for teachers is 3.71—which is, 19% relatively lower.

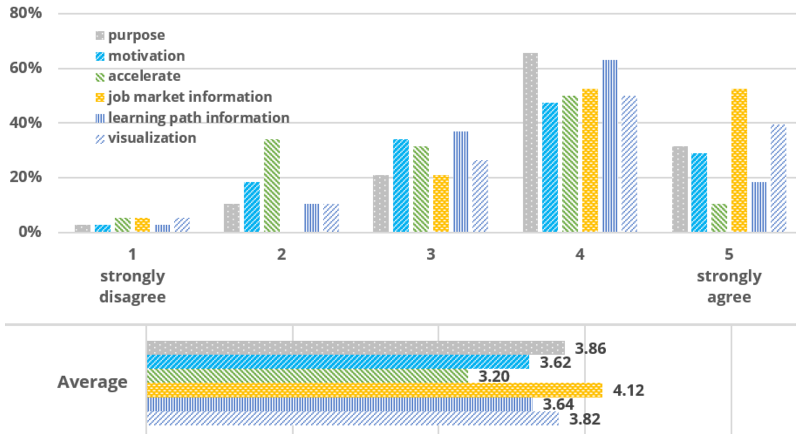

Fig. 8: Teachers' feedback on the *Learning Content-JobMarket Report for Students*.

6.1.2 Evaluation of the *Job-LearningPath Report for Students*

Figure 9 and 10 show the students' and teachers' feedback on the *Job-LearningPath Report for Students* with regard to the criteria *purpose*, *motivation*, *acceleration*, *job market information*, *learning path information* and *visualization*. With averages of 4.47 for both *purpose* and *motivation*, 4.37 for *acceleration*, 4.66 for *job market information*, 4.55 for *learning path information* and 4.79 for *visualization*, the majority of students agrees and even strongly agrees that our report is appropriate for the criteria evaluated.

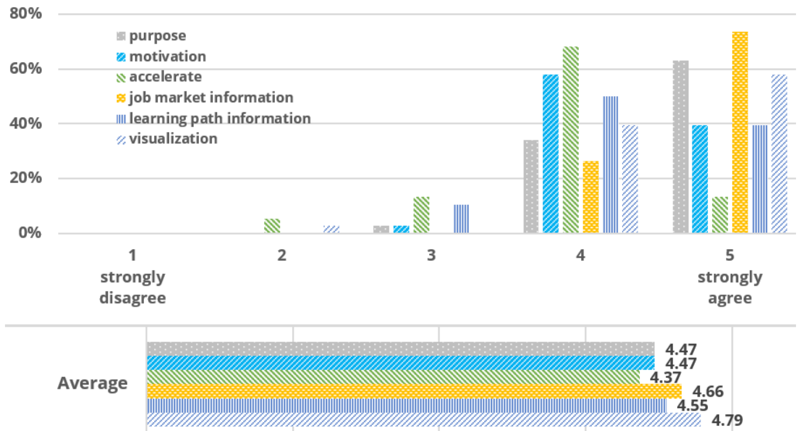

Fig. 9: Students' feedback on the *Job-LearningPath Report for Students*.

Looking at the teachers' feedback on the *Job-LearningPath Report for Students* in Figure 10, we see that with averages of 3.94 for *purpose*, 3.84 for *motivation*, 3.56 for *acceleration*, 3.78 for *job market information*, 3.62 for *learning path information* and 3.84 for *visualization*, the majority of teachers agrees that our report is appropriate for students regarding our evaluation criteria. While the average feedback of the students over all 6 criteria is 4.55, the average of the teachers is 3.76—i.e. 21% relatively lower.

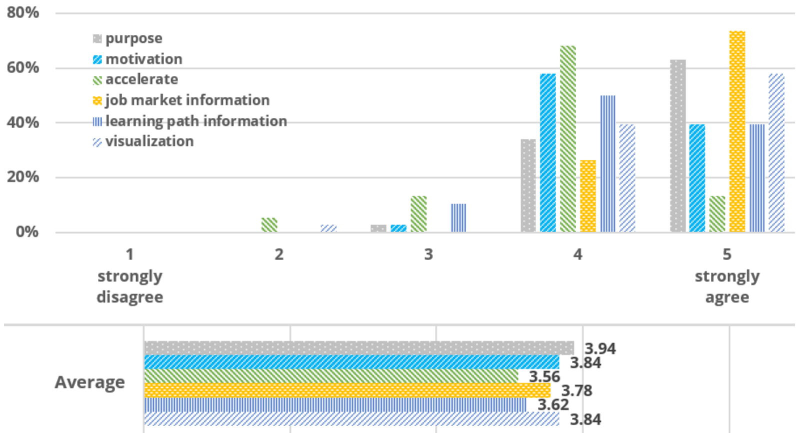

Fig. 10: Teachers' feedback on the *Job-LearningPath Report for Students*.

6.1.3 Evaluation of the *LearningContent-JobMarket Report for Teachers*

Figure 11 illustrates the teachers' feedback on the *LearningContent-JobMarket Report for Teachers* with regard to the criteria *skill importance*, *missing skills*, *acceleration*, *advise*, *targeted education*, and *visualization*. We learn that with averages of 4.20 for *skill importance*, 4.00 for *missing skills*, 3.82 for *acceleration*, 4.12 for *advise*, 4.00 for *targeted education* and 4.12 for *visualization*, the majority of teachers agrees that our report is appropriate for our evaluation criteria. The average feedback of the teachers over all 6 criteria is 4.04, which is 9% and 7% relative better of how they evaluate the *LearningContent-JobMarket Report for Students* and *Job-LearningPath Report for Students*.

7 Conclusion and Future Work

In this work, we have reported how Sentence-BERT [1], K-Means [2], DBSCAN [3], TF-IDF [4] and SVMs [5] are used to build a predictive model capable of identifying job market skills within learning material with an accuracy of 94.2% and an F-score of 86.9%. We presented reports for students that are understandable, show the learning material's purpose, motivate, accelerate the learning, as well as provide job market and learning path information. We also presented a report for teachers that is understandable, shows the importance of present/missing skills in the learning material, accelerates and specifies the learning material creation, and helps them advise students.

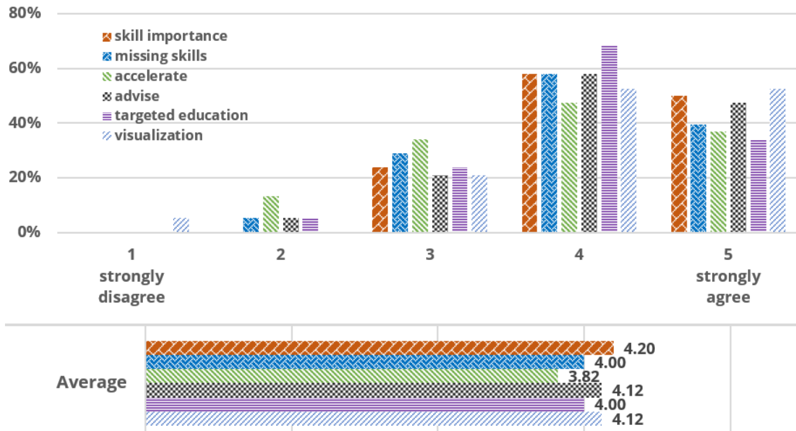

Fig. 11: Teachers' feedback on the *LearningContent-JobMarket Report for Teachers*.

Our future work will focus on further optimizing our pipeline. This will include web scraping more than 100 job postings per job to obtain a more accurate representation of the job market. Additionally, we aim to expand our analysis beyond the six jobs included in our current work. We also plan on including more sections of learning content from a wider range of course books to obtain a larger labeled data set. With a larger labeled data set, we will investigate potential improvements to our *present/missing skills model* using hyperparameter tuning.

References

[1] Reimers, N., Gurevych, I.: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In: EMNLP-IJCNLP, Hong Kong, China (2019)

[2] Steinley, D.: K-means clustering: A half-century synthesis. The British Psychological Society (2010)

[3] Ester, M., Kriegel, H.-P., Sander, J., Xu, X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. KDD (1996)

[4] Sparck Jones, K.: A Statistical Interpretation of Term Specificity and its Application in Retrieval. Journal of Documentation **28**(1), 11–21 (1972)

[5] Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B.: Support Vector Machines. IEEE Intelligent Systems and their Applications **13**(4), 18–28 (1998)

[6] Elfert, M.: Lifelong Learning in Sustainable Development Goal 4: What Does It Mean for UNESCO's Rights-Based Approach to Adult Learning and Education? Int. Review of Education (2019)

[7] Hane-Weijman, E.: Skill Matching and Mismatching: Labour Market Trajectories

of Redundant Manufacturing Workers. Geografiska Annaler: Series B, Human Geography (2021)

[8] Fossen, F., Sorgner, A.: Mapping the Future of Occupations: Transformative and Destructive Effects of New Digital Technologies on Jobs. Foresight and STI Governance (2019)

[9] Rismani, S., Moon, A.: What Does It Mean to be an AI Ethicist: An Ontology of Existing Roles. Computers and Society (2022)

[10] Price, W., Gerke, S., Cohen, I.G.: Potential Liability for Physicians using Artificial Intelligence. JAMA (2019)

[11] Wang, W., Siau, K.: Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Hu-manity: A Review and Research Agenda. Journal of Database Management (2019)

[12] Paranjape, K., Schinkel, M., Panday, R.N., Car, J., Nanayakkara, P.: Introducing Artificial Intelligence Training in Medical Education. JMIR Publications (2019)

[13] Fuller, J., Langer, C., Sigelman, M.: Skills-Based Hiring Is on the Rise. Harvard Business Review (2022)

[14] Sugiyanto, S., Mohammad, B., Herpratiwi, H.: The Influence of Learning Motivation on the Learning Outcomes of Vocational Students at Lampung University. International Journal of Advanced Science and Technology. (2020)

[15] Frymier, A.B., Shulman, G.M.: "What's In It For m Me?": Increasing Content Relevance to Enhance Students' Motivation. Communication Education (1995)

[16] Howard, J., Bureau, J., Guay, F.: Student Motivation and Associated Outcomes: A Meta-Analysis from Self-Determination Theory. Journals.Sagepub.Com (2021)

[17] Azman, A., Simatupang, W., Karudin, A., Dakhi, O.: Link and Match Policy in Vocational Education to Address the Problem of Unemployment. Int. Journal of Multi Science (2020)

[18] Kehdinga, D., Fomunyam, G.: Education and the Fourth Industrial Revolution: Challenges and Possibilities for Engineering Education. Int. Journal of Mechanical Engineering and Technology (2019)

[19] Lauder, H., Mayhew, K.: Higher Education and the Labour Market: An Introduction. Oxford Review of Education (2020)

[20] Li, L.: Education Supply Chain in the Era of Industry 4.0. Systems Research and Behavioral Science (2020)

[21] Bothmer, K., Schlippe, T.: Investigating Natural Language Processing Techniques

for a Recommendation System to Support Employers, Job Seekers and Educational Institutions. In: The 23rd International Conference on Artificial Intelligence in Education (AIED 2022) (2022)

[22] Bothmer, K., Schlippe, T.: Skill Scanner: Connecting and Supporting Employers, Job Seekers and Educational Institutions with an AI-Based Recommendation System. The Learning Ideas Conference 2022 (15th annual conference), Innovative Approaches to Technology-Enhanced Learning for the Workplace and Higher Education (2022)

[23] Bothmer, K., Schlippe, T.: Skill Scanner: An AI-Based Recommendation System for Employers, Job Seekers and Educational Institutions. International Journal of Advanced Corporate Learning (iJAC) **16**(1) (2023)

[24] Konstantinidis, I., Maragoudakis, M., Magnisalis, I., Berberidis, C., Peristeras, V.: Knowledge-Driven Unsupervised Skills Extraction for Graph-based Talent Matching. SETN (2022)

[25] Chaudhari, Y., Jadhav, P.: An End to End Solution For Automated Hiring. IEEE (2022)

[26] Rong, X.: word2vec Parameter Learning Explained. Computation and Language (2014)

[27] Pendyala, V., Atrey, N.: Enhanced Algorithmic Job Matching based on a Comprehensive Candidate Profile using NLP and Machine Learning. IEEE (2022)

[28] Sridevi, G.M., Kamala Suganthi, S.: Artificial Intelligence-Based Job Applicant Profile Quality Assessment and Relevance Ranking Using Clusters in Talent Acquisition Process. Applications of Artificial Intelligence in Engineering (2021)

[29] Dyulicheva, Y., Vernadsky, V.I., Dyulicheva, Y.Y., Bilashova, E.A.: Learning Analytics of MOOCs based on Natural Language Processing. CEUR (2022)

[30] Dogan, Y., Dalkılıç, F., Kut, A., Kara, K., Takazoğlu, U.: Discovering the Same Job Ads Expressed with the Different Sentences by Using Hybrid Clustering Algorithms. Int. Journal of Applied Mathematics Electronics and Computers (2020)

[31] ScrapeOps: Web Scraper. https://scrapeops.io. Accessed on March 1, 2023 (2023)

[32] Stolarek, I., Samelak, A., Figlerowicz, M., Jackowiak, p.: Dimensionality reduction by UMAP for visualizing and aiding in classification of imaging flow cytometry data. Iscience (2022)

[33] Patel, J.: Web Scraping in Python Using Beautiful Soup Library. Getting Structured Data from the Internet (2020)

[34] Chauhan, U., Shah, S., Shiroya, D., Solanki, D., Patel, Z., Bhatia, J., Tanwar, S., Sharma, R., Marina, V., Raboaca, M.: Modeling Topics in DFA-Based Lemmatized Gujarati Text. Sensors (2023)

[35] Narendra, G., Hashwanth, S.: Named Entity Recognition based Resume Parser and Summarizer. International Journal of Advanced Research in Science, Communication and Technology (2022)