Web Platform for Job Recommendation Based on Machine Learning

Iuliana Marin¹ and Hanoosh Amel²

¹Faculty of Engineering in Foreign Languages, University Politehnica of Bucharest, Splaiul Independenței 313, Bucharest, Romania ²Ministry of Education, Directorate of Almuthanna Education, Muthanna, Iraq

Keywords: Jobs, Skills, Recruitment Platform, Recommendations, Machine Learning.

Abstract:

After three years of dealing with a global medical catastrophe, our society is attempting to re-establish normalcy. While companies are still struggling to get back on track, workers have grown afraid to seek new jobs, either because they offer low pay or an uncertain schedule. The result is a disconnected environment that does not merge, even though it appears to. The proposed approach creates a suitable recommender system for those looking for jobs in data science. The first-hand information is gathered by collecting Indeed.com's data science job listings, analysing the top talents that employers value, and generating job ideas by matching a user's skills to openings that have been listed. This process of job suggestion would assist the user in concentrating on the positions where he has the greatest chance of succeeding rather than applying to every position in the system. With the aid of this recommendation system, a recruiter's burden would be decreased because it lowers the quantity of undesirable prospects.

INTRODUCTION

During the pandemic, many organizations urged their employees to work remotely when governments around the world asked enterprises to suspend operations. Many other organizations, on the other hand, began to decrease their operational costs by firing permanent and contract personnel. Individuals who lost their jobs because of the closure are more or less forced to look for new opportunities. This results in a continuous hiring cycle. Therefore, the pandemic became a turning point in employee upskilling and reskilling (Li, 2022).

LinkedIn developed the Career Explorer tool in 2020 to assist laid-off workers to locate possible career transitions based on their abilities (Davis et al., 2020). The tool mapped available applicant skills and identified additional skills candidates could learn to change occupations.

Job seekers have access to various job boards that help in the hiring cycle (e.g., LinkedIn, Glassdoor, Indeed, CareerBuilder). A job seeker searches for a position that appears to be a good fit for him, creates his CV, and applies for it. Given the numerous job boards available, a job seeker will seek out a tool that offers the best features, such as a user-friendly interface, the ability to construct a CV that includes his skills, and the ability to create a user profile. Most of them tend to search for a job that primarily matches his skills, but companies find it challenging to filter candidates.

Instead of a college degree, skills-based hiring requires specific abilities and competencies. It broadens the talent pool, while also narrowing the emphasis by providing more clarity on what is actually needed and wanted from the organization's next great employee, like project management professionals (Dascalu et al., 2015). Skills-based hiring focuses on a candidate's abilities. It eliminates benchmarks such as a four-year degree or a particular number of years of experience, replacing them with skills and competencies obtained in the classroom or on the work.

The aim of the current study was to develop an appropriate recommender system for those persons who look for work in the field of data science. The firsthand information is obtained by scraping data science jobs from the website Indeed.com, analyze top skills required by companies, and generate job suggestions by matching skills from the user's résumé

^a https://orcid.org/0000-0002-7508-1429

to posted opportunities. The objectives of the current research were to scrape job listings from Indeed.com that are generated after typing "data analyst/engineer/analyst" in the input field for job title, keyword, or company. Secondly, another objective was to tokenize and extract keywords for skills from job descriptions, followed by the action to tokenize and extract keywords for skills from the résumé.

The next step was to calculate similarity of keywords from posted jobs and the résumé. An integration of the recommendation process into a web framework is also performed. Another aspect taken into consideration was the design the application, such that a potential user can interact with it. The system generates top 10 job listings tailored to the user's skills stated in his résumé.

The paper is divided into 7 chapters, and each chapter is described as follows: chapter 1 includes the research introduction, objectives, and the motivation for writing the current paper. Chapter 2 contains the description of the state-of-the-art, which includes the theoretical foundation of job recommendation, the setting in which the paper was developed, and a list of similar applications already available on the market. Chapter 3 includes the research methodology utilized to determine the web application requirements. Chapter 4 outlines the presentation of the proposed application with its main functionalities and how they are implemented. Chapter 5 includes the technology and methods employed. The last chapter contains the conclusions and further improvements.

2 RELATED WORK

A recommender system (RS) analyses user preferences and offers them a variety of service options based on their requirements. First, there is a need to distinguish between the roles of the RS on behalf of the service provider and the RS user's role.

If a travel agency or a destination management organization wants to increase its revenue, for example, by selling more hotel rooms or attracting more tourists to the destination, a travel recommender system is implemented in order to satisfy this requirement of the software system (Ravi and Vairavasundaram, 2016). The users' key objectives for using the two platforms are to find a suitable lodging and intriguing events or attractions. As a result, an RS must strike a balance between the needs of these two parties and provide a service that is beneficial to both.

2.1 Algorithms Used in Recommender Systems

There are so many different data and knowledge sources available to RSs, such that the recommendation approach ultimately determines whether they may be employed. Four categories are distinguished by the recommendation algorithm: content-based filtering, collaborative filtering, rule-based methods, and hybrid approaches (Afoudi et al., 2021; Wayissa et al., 2022).

Content-based Filtering (CBF), unlike collaborative filtering, which chooses things based on the correlation between users with similar preferences, CBF chooses items based on the correlation between the items' content and the user's preferences (Fkih, 2022). The user is assumed to select items with similar qualities. Because user profiles are based on a characteristic of the prior item selected by the user, the researchers (Ko et al., 2022) claim that the filtering approach has a propensity to over-specialize when proposing an item to a user profile.

However, due to the nature of the employment market, the position placed on the job board will only be open for a short period of time. When it comes to leisure, the user tastes might change for several reasons, but when it comes to work, users prefer to look for jobs that let them put their abilities to use. New job recommendations can be given when a user's choices change, such as when he decides to update his job domain by adding his new abilities and if he so chooses.

Collaborative filtering (CF) takes advantage of users' preferences for prior favourites of comparable items (Chi Yi and Kang, 2021). It is one of the best ways for recommendations, and interest in it from academia and business is growing. Collaborative filtering is only partially successful in some application areas due to the cold-start problem, which happens when historical data is too sparse (also known as the sparsity problem) or when new users have not rated enough items, or both. Even though content-based filtering approaches are outperformed by collaborative filtering, neither a job, nor a similarity matrix can be developed because of the nature of the hiring process.

In rule-based filtering (RF) systems, consumers are given recommendations based on manually or automatically generated decision rules. Many websites that currently make use of personalisation or recommendation technologies employ manual rule-based methods, which is not the case of the proposed platform that is fully autonomous.

Recommender systems provide website owners the ability to create rules, many of which are based on the demographic, psychographic, or other individual characteristics of visitors (Konstantakis et al., 2022). The primary drawbacks of RF solutions are the techniques employed to create user profiles. The input is generally skewed since it is a subjective description of the interests of users or the interests of the users themselves. Furthermore, system performance declines over time as the profiles get older since they are typically stagnant.

A combination of several suggesting techniques generates a hybrid recommender system. When compared to collaborative or content-based systems, hybrid recommender systems often provide more accurate recommendations (Deschênes, 2020). This is due to ignorance of the domain dependencies of collaborative filtering and user preferences in a content-based system.

When ranking the results of a query, search engines consider textual similarity. Information retrieval using a vector model is one of the text similarity's most significant uses (Christino et al., 2022). Documents are sorted in this type of application based on how pertinent they are to an input query. The two methods that may be used to quantify the degree of similarity between two texts are lexical and semantic similarity.

A sequence of strings that are related to one another can be used to determine how comparable a string's lexical similarity is. When calculating a word's semantic similarity, the context of the term is considered. The degree of resemblance may be assessed using the Jaccard and Cosine similarity metrics (Pernisch et al., 2021).

2.2 Models of Successful Job Recommender Systems

A recommendation engine can be added to a website. Google is one of the most familiar with website which employs its Google Advertising recommender system to show relevant ads.

According to various research on the topic, the LinkedIn recommender uses content matching and collaborative filtering to identify businesses or jobs that a user might be interested in. The key elements of recommender systems are the users' jobs, education, summaries, specializations, experiences, and skills from information on their LinkedIn profiles. Data regarding a member's relationships, affiliations with organizations, and companies they have followed, for instance, are obtained through their activity.

To correctly match members to jobs, LinkedIn uses "Entity Resolution," which is the process of separating apart appearances of real-world entities in different records or references. In LinkedIn's entity resolution process, which makes use of many standards for business standardization, machine-learned classifiers are employed (Urdaneta-Ponte et al., 2022).

Given that a college degree is not necessarily necessary for professional success, Indeed, which attracts more than 250 million unique visitors each month, aims to provide goods that open doors for everyone seeking for work. The free services offered by Indeed allow job searchers to look for employment. Users may register, add their resume, and seek for positions that suit their requirements. To create its recommendation engine, Indeed started with an Apache Mahout MVP and then switched to a hybrid offline/online pipeline (Alsaif et al., 2022). Algorithms, system architecture, and model format were gradually improved along the way to solve a variety of problems.

The usage of a recommendation engine for a web platform addressed to students and people with IT skills, is important. In Romania, platforms like ejobs, bestjobs, hipo, cvjobs, jobzz, or the ones from Iraq, such as Bayt, Hawa, do not involve scraping jobs from another website, like in the current paper, where Indeed.com is used. Many remote jobs appeared during the COVID-19 pandemic and students can benefit from it to gain experience and skills.

3 METHODOLOGY

Twenty respondents with a range of educational backgrounds including business, technical, legal, communication, and marketing degrees completed an online survey. From their replies, firstly, most people use social networks or job boards to find their present employment. Secondly, the majority of people want to work in a position that advances their professional and personal objectives, therefore they seek employment that is suited to their background, abilities, and interests.

Every respondent agreed that talents are more important than a college degree. The majority of those surveyed said they would utilize a job board that creates job advertisements based on their qualifications. The functional and non-functional requirements for the current research have been identified after assessing the replies.

3.1 Functional Requirements

Several functional requirements have been identified as an outcome of the survey. The first one is to create a user account, such that the user should have the option to register. The second requirement is to manage login and logout of users. Another requirement is to upload a CV, such that the user should be able to upload a PDF version of his resume.

Update CV is another requirement which is needed if the user chooses to erase his previous entry, he should be allowed to add a new CV. Delete CV is for the situation when the user needs to have the option of deleting his CV. Following the addition of the user's CV and/or chosen location, the user should be able to conduct a job listing search. Moreover, the user should be able to add the location where he wants to look for a certain job.

3.2 Non-Functional Requirements

Usability, Correctness metrics, response time and a friendly user interface are the non-functional requirements of the proposed system. Regarding usability, even non-technical users should find the website easy to use. The average user decides whether to stay on a website after only 0.05 seconds. It must also be easy to use because it is not a job board, but rather a tool that will help with the job search process.

Correctness measures, including recall, accuracy, and precision are needed for recommendation systems standards. In what concerns response time and performance, in many cases, the application's responsiveness is a crucial consideration, sometimes even more so than the accuracy of the conclusions. When the number of suggestions needed each time unit is known, a better selection of algorithms may be made. For the user-friendly design is needed an intuitive user interface that is uncluttered of distracting images.

4 PROPOSED JOB RECOMMENDATION SYSTEM

The hiring suggestion system was developed in response to the need from job seekers for a skill-based hiring recommendation. It aims to replace conventional demands like a four-year degree or a set number of years of experience with abilities and competences acquired in the classroom or on the work. Currently, in Romania and Iraq, there is no such similar platform available.

The provided tool was created with the aid of Python, Flask Framework, Firebase, HTML, Bootstrap, making it ideal for the demands of a modern hiring process assistance. The system may be accessed over an Internet connection using any web browser on any device. It is a multi-tiered web application, as illustrated in Figure 1. Its intuitive user interface adheres to responsive web design.

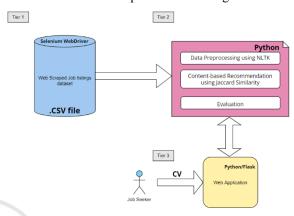


Figure 1: Application three tier design.

The landing page (Figure 2) is where the user will have their initial interaction with the web application. Depending on what he needs, he can be taken from this page to either the Sign Up form or the Log In form, if he already has an existing account.

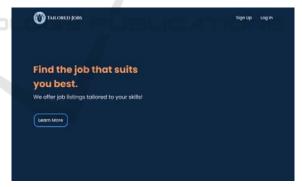


Figure 2: Landing page.

The following step of the user experience within the online application is a login or registration (Figure 3). Each user will be required to set up an account in order for the information to be saved in a manner that is specific to that user. If the user has already been registered with the site or if the login credentials are entered incorrectly, an error message will be displayed.



Figure 3: Sign-up form.

Next, after the user introduces his credentials, he will be redirected to the page where he will upload his CV in PDF format and fill in the location where he wants the application to search for job listings (Figure 4). Moreover, if he changes his mind and does not want the same CV to be uploaded, he can delete it and introduce a new one. After all the necessary information has been introduced, the user will click on the "Generate jobs" button and will be redirected to the "Job Listings" page. Moreover, he can log out at any time.

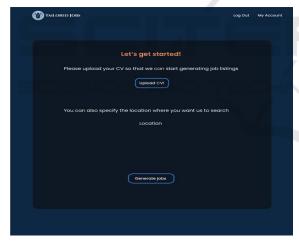


Figure 4: Job generation page.

In the "Job Listings" page (Figure 5), the user will be provided with top 10 job listings tailored to his skills. Here, for each job listing, he can see the company name, job title and location. If he wants to see more information, he can click on the desired job listing and will be redirected to the "Job Details" page.



Figure 5: Job listings page.

In the "Job Details" page (Figure 6), we have more detailed information such as the location of the job vacancy, the type of job, industry, salary and the description. Moreover, if the user finds the posting appealing, he can apply to it by clicking on the "Apply" button that will redirect him to the initial Indeed job posting.

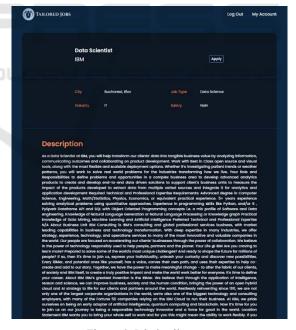


Figure 6: Job details page.

5 TECHNOLOGICAL CHOICES

There were needed a number of tools to extract data from Indeed.com. The unstructured nature of the text

data necessitates pre-processing. For each position we are considering, we tokenize the job description, remove terms from the NLTK stop words list, and then filter on a list of common data science-related skill words. If a position is appropriate for a job seeker, it may be determined by comparing the skill keywords in a sample CV and a job description. Once a score has been computed, the top 10 works will be shown on the user's dashboard. Below is a detailed description of the feature development and methods used in this project.

Web scraping, also known as data scraping or web data extraction, is a method that utilizes automation to collect data from websites. The automated programs, which may be referred to as bots or crawlers, navigate to and interact with a large number of web pages. They then extract useful information from those pages, parse it, and save it in structured data formats that are compatible with software such as spreadsheets, databases, and analytical tools.

The automated program that was used for this project is called Selenium. Selenium was first made as a tool to test how websites work, but it quickly became a general tool for automating web browsers that is used for web-scraping and other tasks. Selenium WebDriver is the first browser automation protocol made by the W3C organization and is a middleware protocol service that sits between the client and the browser and translates commands from the client into actions for the browser (Garcia et al., 2020). With the help of the WebDriver, it was managed to scrape data science jobs from indeed.com, gathering data scientist/engineer/analyst jobs posted in the last 30 days in 5 major Romanian cities, i.e., Bucharest, Iasi, Cluj-Napoca, Constanța and Timişoara. A JSON file was created to store the results (Figure 7).

```
m = random.randint(1,5)
time.sleep(m)
# Retrieve single job url
link = job_lk['job_link']
driver.ger(link)
# Job city and province
location = job_lk['location']
# Job city and province
location = job_lk['location']
# Job city and province
location = job_lk['location']
# Job city and province
| Job city and provi
```

Figure 7: Excerpt of code used to retrieve information from Indeed listings.

Research in this study relies heavily on Natural Language Processing (NLP). Iterating over each job

description, it was tokenized, cleaned it up by removing stop words, and then filtered it using a list of common data science-related skills. Stop Words: A stop word is a regularly used term (such as "the," "a," "an," and "in") that a search engine has been configured to ignore. Python's Natural Language Toolkit (NLTK) includes a library that stores a list of stop words in 16 different languages. They are located in the nltk data directory.

Job descriptions have been filtered by using a list of data science related skills, which became an overall dictionary (Figure 8).

Figure 8: Dictionary of data science skills.

PyPDF2 is a pure-python PDF library that is both free and open-source. It is able to split, merge, crop, and otherwise change the pages of PDF files. PDF files can also have user-specific data, viewing choices, and password protection added using this tool. PyPDF2 has the ability to extract text as well as metadata from PDF files. Using the PyPDF2 python tool, keywords detected in the overall dictionary from résumés.

Job recommendations are based on the similarity of skill keywords in the job description and the résumé. The CV is automatically analysed, also based on NLP, as in the case of jobs. To perform the match between jobs and a candidate, the current research employed the Jaccard similarity (i.e., intersection over union of two groups). In this case, more matching keywords and fewer mismatched keywords lead to higher scores (between 0 and 1).

For example, the following are my top five job matches in Bucharest, Ilfov, when calculating the similarity between the skill keywords from the résumé and the skill keywords from job descriptions, as in Figure 9.

link	location	title	company	salary	desc	keywords	similarity
https://ro.indeed.	Bucharest, Ilfov	Data Analyst	Lugera & Makler	Cu norma	Detaliile	['r', 'python', 'm	0.33333
https://ro.indeed.	Bucharest, Ilfov	Data Analyst (m/f/d)	METRO.digital	NaN	Today	['sql', 'r', 'tablea	0.33333
https://ro.indeed.	Bucharest, Ilfov	Senior Data Analyst, Central Operations	Bolt	NaN	<gh-< td=""><td>['sql', 'bachelor'</td><td>0.25</td></gh-<>	['sql', 'bachelor'	0.25
https://ro.indeed.	Bucharest, Ilfov	Data Scientist	SAS Institute Inc	NaN	:	['r', 'sas', 'pytho	0.25
	0 1 . 110	0 1 0 1 11 1	10.44		m . Int		0.0

Figure 9: Top 5 job listings generated for Bucharest, Ilfov.

6 CONCLUSIONS AND FUTURE WORK

To determine the degree to which an available position and its user are similar, the current research on recommender systems in the hiring industry looks at what abilities are necessary for each job. On the other hand, the entertainment industry's recommender system relies on user input. A user rates a particular item, and this rating is used to produce an item recommendation to a user. But this concept of forecasting the likelihood of a user to choose an item would be inaccurate when viewed from the perspective of a job seeker.

In this study, it was employed a content-based filtering to recommend a job that is similar to the user's provided information which is automatically analysed. Instead of applying to all the jobs in the system, this procedure of recommendation would help the user focus on the ones that he is most likely to succeed at. A recruiter's workload would be reduced with the help of this recommendation system because it reduces the number of unsuitable candidates. Currently, there is no such similar solution in Romania and Iraq. Students from the IT domain will be encouraged to find a job easily and even work remotely, as more and more such offers appeared available after the COVID-19 pandemic emergence. Nevertheless, students can find part-time job offers on their faculty premises. This is essential for the students who need to support themselves during their studies. Women will also be helped to find a job and adapt in a progressive world, based on their religious and cultural constraints.

Concerning the recommendation system, for future work we will construct a data skill vocabulary (e.g., IT knowledge, vocabulary, and industry jargon) by exploring job descriptions rather than using a predefined collection of words. Also, there will be a need to undertake additional research on content-based filtering and other filtering techniques from the point of view of a job seeker.

Concerning the web application, additional functions that can optimize the flow may be included as part of subsequent enhancements to the platform. These functions might include a detailed User Profile, in which the user is able to view the job advertisements that he has marked as favorites; a Company Profile, in which a possible recruiter is able to view the User Profile of a potential candidate, and real-time private chat rooms.

REFERENCES

- Afoudi, Y., Lazaar, M., Al Achhab, M. (2021). Hybrid Recommendation System Combined Content-Based Filtering and Collaborative Prediction using Artificial Neural Network. In *Simulation Modelling Practice and Theory*, 113, 1-10.
- Alsaif, S. A., Hidri, M. S., Ferjani, I., Eleraky, H. A., Hidri, A. (2022). NLP-Based Bi-Directional Recommendation System: Towards Recommending Jobs to Job Seekers and Resumes to Recruiters. In *Big Data and Cognitive Computing*, 6(4), 1-17.
- Chi Yi, A. L., Kang, D.-K. (2021). Experimental Analysis of Friend-And-Native Based Location Awareness for Accurate Collaborative Filtering. In *Applied Sciences*, 11(6), 1-17.
- Christino, L., Ferreira, M. D., Paulovich, F. V. (2022). Q4EDA: A Novel Strategy for Textual Information Retrieval Based on User Interactions with Visual Representations of Time Series. In *Information*, 13(8), 1-24.
- Dascalu, M.-I., Bodea, C.-N., Marin, I. (2015). Semantic Formative E-Assessment for Project Management Professionals. In 2015 4th Eastern European Regional Conference on the Engineering of Computer Based Systems, 1-8.
- Davis, J., Wolff, H.-G., Forret, M. L., Sullivan, S. E. (2020). Networking via LinkedIn: An Examination of Usage and Career Benefits. In *Journal of Vocational Behavior*, 118, 1-15.
- Deschênes, M. (2020). Recommender Systems to Support Learners' Agency in a Learning Context: A Systematic Review. In *International Journal of Educational Technology in Higher Education*, 17, 1-23.
- Fkih, F. (2022). Similarity Measures for Collaborative Filtering-Based Recommender Systems: Review and Experimental Comparison. In *Journal of King Saud University Computer and Information Sciences*, 34(9), 7645-7669.
- Garcia, B., Gallego, M., Gortazar, F., Munoz-Organero, M. (2020). A Survey of the Selenium Ecosystem. In *Electronics*.
- Ko, H., Lee, S., Park, Y., Choi, A. (2022). A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. In *Electronics*, 9(7), 1-29.
- Konstantakis, M., Christodoulou, Y., Aliprantis, J., Caridakis, G. (2022). ACUX Recommender: A Mobile Recommendation System for Multi-Profile Cultural Visitors Based on Visiting Preferences Classification. In *Big Data and Cognitive Computing*, *6*(4), 1-11.
- Li, L. (2022). Reskilling and Upskilling the Future-ready Workforce for Industry 4.0 and Beyond. In *Information Systems Frontiers*, 1-16.
- Pernisch, R., Dell'Anglio, D., Bernstein, A. (2021). Toward Measuring the Resemblance of Embedding Models for Evolving Ontologies. In *K-CAP'21: Knowledge* Capture Conference, 177-184.
- Ravi, L., Vairavasundaram, S. (2016). A Collaborative Location Based Travel Recommendation System

- through Enhanced Rating Prediction for the Group of Users. In *Computational Intelligence and Neuroscience*, 1-29.
- Urdaneta-Ponte, M. C., Oleagordia-Ruiz, I., Mendez-Zorrilla, A. (2022). Using LinkedIn Endorsements to Reinforce an Ontology and Machine Learning-Based Recommender System to Improve Professional Skills. In *Electronics*, 11(8), 1-19.
- Wayissa, F., Leranso, M., Asefa, G., Kedir, A., Salau, A. O. (2022). Pattern-Based Hybrid Book Recommendation System using Semantic Relationships. In *ResearchSquare*, 13, 1-12.

