Journal of Information Systems Engineering and Management
2025, 10(8s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

BERT-based Job Recommendation System Using LinkedIn Dataset

Liala Almalki[1], Norah Al-Alsheikh[2], Fatmah Bahareth[3], Manal Abdullah[4], Roaa Alsahafi[5]

[1]*Department of Information Systems*
King Abdulaziz University
Jeddah, Saudi Arabia
lalmalki0081@stu.kau.edu.sa
[2]*Department of Information Technology*
Imam Mohammad Ibn Saud Islamic University
Riyadh, Saudi Arabia
naalshikh@imamu.edu.sa
[3]*Department of Information Systems*
King Abdulaziz University
Jeddah, Saudi Arabia
fbahareth@stu.kau.edu.sa
[4]*Department of Information Systems*
King Abdulaziz University
Jeddah, Saudi Arabia
maaabdullah@kau.edu.sa
[5]*Department of Information Systems*
King Abdulaziz University
Jeddah, Saudi Arabia
ralsahafi0063@stu.kau.edu.sa

ARTICLE INFO

Received: 20 Oct 2024

Revised: 12 Dec 2024

Accepted: 24 Dec 2024

ABSTRACT

Amidst rapid technological progress, bridging the gap between job seekers and employers has become increasingly important as the job market evolves. This research presents a job recommendation system that leverages the Bidirectional Encoder Representations from Transformers (BERT) model, a powerful Natural Language Processing (NLP) framework. The system ensures precise and personalized recommendations by understanding the semantic relationships within job descriptions and user profiles. The model integrates contextual matching of skills and preferences, addressing the limitations of traditional content-based methods. Evaluation results demonstrate effectiveness in recommendation accuracy. This research proposes an alternative approach to the job-matching process by harnessing BERT's bidirectional contextual capabilities. The presented work has significant implications for human resources platforms, job portals, and recruitment solutions in an increasingly digital workforce ecosystem.

Keywords: Recommendation systems, Job recommendations, BERT model, Content-Based filtering

INTRODUCTION

The job market has become more challenging for job seekers, from experienced workers to recently graduated pupils [1]. There is a significant disparity between the skill sets of candidates and the job specifications they are applying for, which leads to increased competitiveness and widespread ambiguity about each application. Coming up with innovative solutions is key to addressing this imbalance.

A pivotal resource in addressing the challenges of the current job market is the emergence of job recommendation systems that leverage advanced algorithms and sophisticated data analytics. These systems empower users to identify positions that closely align with their unique skill sets, professional experiences, and career aspirations by offering personalized job suggestions [2]. By curating a more tailored selection of opportunities, job recommender systems not only enhance recruiting efficiency for employers but also streamline the job-seeking process for candidates.

Many companies are increasingly adopting Digital Human Resource Management Systems that are integrated with professional networking platforms like LinkedIn. LinkedIn has firmly established its reputation as a premier resource for job opportunities, brand connections, and professional growth, among social media platforms on the market today. Millions of users share profiles, connect with others, post, or search for jobs. This integration facilitates the digital collection, storage, and processing of talent-related data [3]. By leveraging LinkedIn's advanced algorithms and rich data insights, companies can enhance their ability to match job postings with the most suitable candidates and identify and address skill gaps within the workforce. This, in turn, leads to streamlined recruitment processes, improved quality of hires, and the strategic development of employees through targeted training programs.

Typically, job recommender systems employ content-based recommendation techniques, which primarily rely on the explicit context and interests derived from user profiles [4,5]. Content-based recommender systems are a form of personalized recommendation system that suggest items to users based on the characteristics of the items' content and the preferences of the users [6].

However, dealing with content in recommender systems poses several challenges [7]. Handling content effectively necessitates an in-depth understanding of the text and the relationships between words and sentences, which traditional linguistic modelling often fails to capture directly. To address this, advanced linguistic models like the BERT model should be utilized, as they excel at revealing contextual meaning and semantic relationships within text. The BERT model (Bidirectional Encoder Representations from Transformers) is a powerful deep learning model developed by Google for natural language processing (NLP) tasks [8]. It leverages the Transformer architecture to understand the word context in a text by considering both their left and right surroundings, making it bidirectional. By leveraging BERT's ability to process bidirectional context and understand nuanced word relationships, more accurate recommendations can be achieved.

The present research aims to propose a job recommendation system that precisely aligns the skills of job seekers with the most relevant LinkedIn job postings by leveraging advanced NLP techniques, including the BERT model for understanding contextual relationships within job descriptions and skills. The paper presents the proposed model's results along with an evaluation of its performance.

This paper is organized in the following manner. Section II provides the background information. Section III reviews related work. Section IV describes the process of data preparation. Section V presents data analysis. Section VI describes the proposed system. Section VII is devoted to results and evaluation. Lastly, Section VIII concludes the paper by highlighting the main points, along with future work.

BACKGROUND

A. Big Data Analytics

The proliferation of social media platforms paves the way for huge amounts of data generated by millions of users every moment. Users of these platforms generate large-scale data through their connections and interactions. User-generated content encompasses posts, reviews, comments, messages, and more. As such, social media has become a significant source of big data, an asset that could be harnessed to gain worthwhile insights into patterns and trends.

Big data refers to vast and complex datasets exceeding the processing capacity of traditional computing systems, coming from various sources, such as transactions, sensors, and social media platforms [9]. Handling big data poses a challenge to traditional data processing approaches. This is due to its massive size, fast generation, and diverse types that traditional infrastructure cannot efficiently handle [10]. The field of big data tackles this challenge and deals with distinct needs like integrating unrelated datasets, managing unstructured data, and extracting hidden information promptly. Many tools are available for processing big data [11]. For example, Hadoop, and Apache Spark. Hadoop is an Apache-developed open-source framework. Its various components are designed to handle enormous amounts of structured, unstructured, and semi-structured data that are dispersed across multiple files and processed concurrently. The MapReduce programming model applies the divide-and-conquer strategy to the Hadoop Distributed File System. Spark is an open-source engine made for sequential data analysis. Many businesses utilize Spark and Hadoop as complimentary platforms to study large data since they may offer visual and predictive features.

Carrying out big data analytics can yield valuable insights into improving operational efficiency, identifying new opportunities, making accurate predictions, and supporting informed decision-making [12], [13]. There are different analytics techniques to help organizations understand their data and make informed decisions [14]. These types are descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analysis. An explanation of each technique is provided below.

- Descriptive analysis uses correlations and patterns to explore and understand data.
- Diagnostic analytics is used when an event needs to be investigated for its cause.
- Predictive analytics uses historical data to estimate future events. This includes specific approaches, most of which are based on statistical techniques, that forecast future outcomes using both historical and present data.
- Prescriptive analytics although very important, this kind of analysis is not often applied. Prescriptive analysis is more equipped to provide specific answers and solutions for particular situations. Descriptive and predictive analysis are both linked to prescriptive analysis.

B. RECOMMENDER SYSTEM

Recommender systems process and filter information to deliver recommendations and personalized outputs aligned with the user's needs and preferences [6]. Recommender systems have evolved since their introduction. The first generation of recommender systems includes collaborative filtering, content-based, and a hybrid combining both approaches. The second generation of recommender systems brought the concept of knowledge about items and users, the context-aware recommender systems. The third generation included psychological behavior and emotional status in the process of decision-making. The basic classification of recommender systems includes 1) collaborative filtering, 2) content-based, 3) hybrid recommender system and 4) Knowledge-based system. Some other recommender systems are also implemented such as the Demographic System and Community-based systems [6]. A collaborative filtering system works on giving user ratings to some items. It does not require information about users or the items, rather it works on ratings. On the other hand, the content-based system requires item descriptions and user preferences for recommendations. The third type is the hybrid recommender system, which combines collaborative filtering and content-based approaches; together they enhance the quality of recommendation. The fourth type is the Knowledge-based system, it does not require user ratings, but it generates recommendations based on user requirements.

In this research, we adopted a primarily content-based approach, focusing on recommending jobs whose characteristics align with user profiles. Our approach leverages advanced NLP techniques, including the BERT model, to understand contextual relationships within job descriptions and user skills, ensuring more accurate and personalized recommendations.

Outlined below is a detailed explanation of the selected approaches.

1) Content-based System

The fundamental equation of Content-Based Filtering involves assessing the similarity between an item and the user's profile, typically employing vector representations of attributes [15]. Each item i is denoted by a feature vector $i = [f1,f2,...,fn]$, where $f1,f2,...,fn$ indicate the properties (features) of the item. A user's preferences are encapsulated in a profile vector $u=[w1,w2,...,wn]$, where $w1,w2,...,wn$ denote the weights reflecting the user's interest in the respective attributes.

The similarity $S\ (i,\ u)$ between an item (i) and a user profile (u) is determined using a similarity metric, such as Cosine Similarity:

$$S(i,u) = \frac{\mathbf{i} \cdot \mathbf{u}}{||\mathbf{i}||\,||\mathbf{u}||} \quad (1)$$

The system calculates the similarity score $S\ (i,\ u)$ for every item in the dataset. Items with the greatest similarity ratings are suggested to the user.

2) BERT Model

BERT is a pre-trained deep learning model designed to generate bidirectional contextual representations from large-scale unlabeled text data [8]. It achieves this by jointly considering both the left and right context of each word at every layer of its architecture. This bidirectional approach allows BERT to understand the full context of a word or phrase in relation to its surrounding text. Without significant task-specific architectural changes, the pre-trained BERT model can be developed as a state-of-the-art model for a wide spectrum of tasks, including language inference and question answering. BERT is empirically strong and theoretically straightforward.

The key features of the BERT model include Pre-training and Fine-tuning. BERT undergoes pre-training on extensive text corpora through two tasks: Masked Language Modeling (MLM), in which words are randomly obscured in sentences, and the model forecasts them. Next Sentence Prediction (NSP), the model forecasts the likelihood that one sentence follows another inside a text. Following pre-training, BERT can be fine-tuned for certain downstream tasks, such as sentiment analysis or question answering, utilizing task-specific labelled data.

Another important feature is Tokenization with WordPieces. BERT tokenizes text into smaller subword units for example, playing becomes play ##ing. This helps handle out-of-vocabulary words effectively.

C. Model Architecture

The BERT architecture has revolutionized NLP field by enabling more accurate and context-aware understanding of text. BERT employs a transformer-based design, using bidirectional context to interpret words in relation to all surrounding words in a sentence, rather than processing them in a unidirectional manner like previous models. It incorporates several key components to process input tokens.

Figure 1 shows the BERT model architecture. Each token sent into BERT is denoted by:

- Token Embeddings: The encoded significance of words or sub-words.
- Segment embeddings: Differentiate between two sentences.
- Position embeddings denote the location of tokens inside the sequence.

BERT employs just the encoder component of Transformer design. Each encoder layer comprises Self-Attention, which identifies connections among tokens inside the sequence, and a Feed-forward network that implements non-linear transformations on self-attention outputs.

RELATED WORK

The recommender system research community has recently developed an interest in job recommendation systems. Several well-known systems and research on this topic are available, including PROSPECT [16]. PROSPECT is a content-based system that assists in recruiting candidates by extracting key aspects of their profiles, such as skills and previous experience.

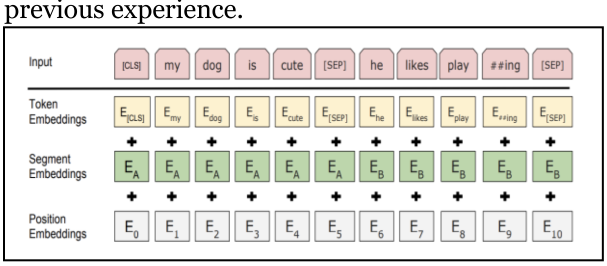

Fig. 1. BERT model architecture

In addition, a research [17] presented a content-based recommendation algorithm that enhances the Minkowski distance to tackle the difficulty of connecting individuals with job opportunities. The suggested approach, FoDRA (Four Dimensions Recommendation Algorithm), assesses a job seeker's fit for a role more dynamically by utilizing a structured representation of both jobs and profiles of candidates, which are derived from a content analysis of unstructured job descriptions and candidate CVs. They did an experimental assessment to evaluate the quality and efficacy of FoDRA. The results indicated that FoDRA yields encouraging outcomes and establishes new opportunities in the domain of Job Recommender Systems.

Furthermore, a study [18] developed a recommendation system that matches user skills and preferences to job listings. Users' input details, such as location, job type, and skills, were pre-processed into a skill profile using TF-IDF. Additionally, job listings were filtered using a Spark SQL query based on user preferences. Cosine similarity was calculated between the user's skill profile and job profiles to rank the matches. Finally, the most relevant jobs were recommended based on the similarity scores. The recommender system evaluated over 10 test cases including different skills and target jobs. The results demonstrated the weaknesses of TF-IDF and cosine similarity, including their inability to capture semantic meaning or closely related words, as well as their sensitivity to noise.

The field of job recommendation has long relied on well-structured data and traditional machine learning techniques, often constrained by significant limitations. However, in the realm of artificial intelligence, the emergence of deep neural networks has revolutionized possibilities. These advanced models represent one of the most transformative technologies of our era, consistently outperforming traditional approaches and achieving unprecedented results across diverse applications. Their capacity to capture complex patterns and deliver highly accurate predictions marks a pivotal shift in how we approach challenges in job recommendations. Many studies have addressed the use of deep learning techniques in the field of job recommendation. For instance, research [19] developed a job recommendation system that integrates web scraping with NLP techniques. The system gathered job data from a leading IT job portal using a customized and tailored web scraping service and then preprocessed the data. It applied the BERT model for generating embeddings for both job and profiles of users. The system generates ten job recommendations, based on user input, by analyzing job descriptions and keywords that are relevant using the Universal Sentence Encoder.

Another study [20] explored two key challenges related to job title polysemy: the use of multiple names for the same position and the use of identical titles for distinct roles. Additionally, the study highlighted the lack of contextual information, which could improve the understanding of job titles and refine prediction accuracy. To address these issues, the research developed a job recommendation system leveraging deep learning methods, specifically the BERT model, for natural language processing. The system demonstrated exceptional efficiency across various tasks, achieving good results and significantly enhancing the precision of job recommendations.

DATA PREPARATION

Data preparation encompasses the processes of cleaning, transforming, and organizing data to ensure it is ready for analysis [21]. This step ensures data quality, consistency, and readiness for getting accurate and efficient recommendations. This section comprises three sub-sections: dataset description, data cleaning, and data preprocessing.

D. Dataset description

The dataset selected is “1.3M LinkedIn Jobs & Skills (2024)” [22]. It encompasses 1.3 million job listings sourced from LinkedIn in 2024. It is applicable for different purposes, including building models for job recommender systems. Figure 2 shows the dataset structure. The dataset encompasses three data files: job_skills, job_summary, and linkedin_job_postings. The dataset encompasses valuable attributes, including the job description, the corresponding required skillsets, positions, levels, work mode, locations, etc.

E. Data Cleaning

The cleaning stage involved the following steps. (1) The job data files were loaded into DataFrames and then the transformation process was applied to the job link feature to transform its value from being a URL to a simplified identifier (2) After that, any existing duplicates were removed from all three data files. (3) To streamline processing, data in the three DataFrames that share a common identifier was integrated, linking relevant information (e.g., job skills, or postings) by the job_link_cleaned key, and ensuring that the final DataFrame contains only records presented in all joined DataFrames. (4) A cleaning pipeline was implemented to standardize job-related data for analysis and to remove unnecessary whitespace, tabs, non-ASCII characters, and markdown formatting from various textual columns. This was to ensure data uniformity and to enhance the quality of inputs for downstream tasks. This was applied across multiple columns in the dataset, including job_title, job_skills, and job_location, ensuring all relevant fields were clean and consistent.

F. Data Preprocessing

The data preprocessing workflow began by (1) loading the CSV into a Spark DataFrame with specific configurations to handle formatting-related issues, such as multiline records,

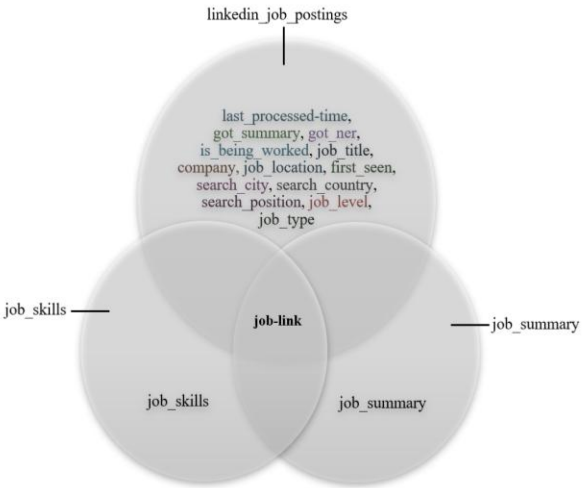

Fig. 2. LinkedIn dataset

escape characters and schema inference. The unwanted line breaks and extra whitespace in text fields, specifically the job_title and job_skills columns, were removed, ensuring clean and consistent formatting. (2) The job_title and job_skills columns were combined into a single column named job_text, with a space separator between the values. The combined column is further trimmed to eliminate any unnecessary spaces. Rows, where the resulting job_text column is empty, were filtered out to ensure data integrity. Lastly, (3) the cleaned and processed DataFrame is previewed, allowing validation and ensuring the data is ready for downstream tasks.

DATA ANALYSIS

Recognizing trends in job and skill demand goes beyond assisting job seekers; it shapes career paths and guides governmental policy decisions. Such awareness is crucial for strategic career planning and staying competitive in an ever-evolving professional landscape. This section focuses on analyzing a dataset to identify trends in job market demands, skill requirements, and the most required jobs. We explored the geographic distribution of job demand. Further, we looked into the most in-demand skills and the companies with the highest job openings on LinkedIn, spanning various industries such as healthcare, retail, and energy.

The results of our analysis showed that the United States has an overwhelmingly high number of job postings compared to other locations, suggesting it is the primary hub for job demand, see Figure 3. Consequently, we focused on exploring the most in-demand jobs in USA. Figure 4 shows that accounts for the largest share at 15.8%, indicating Lead Sales Associate - FT role is the most in-demand job in the U.S. Lead Sales Associate - PT comes with 10.6%, suggesting significant demand for both full-time and part-time positions in this role. Shift Manager represents 11.7% of demand,

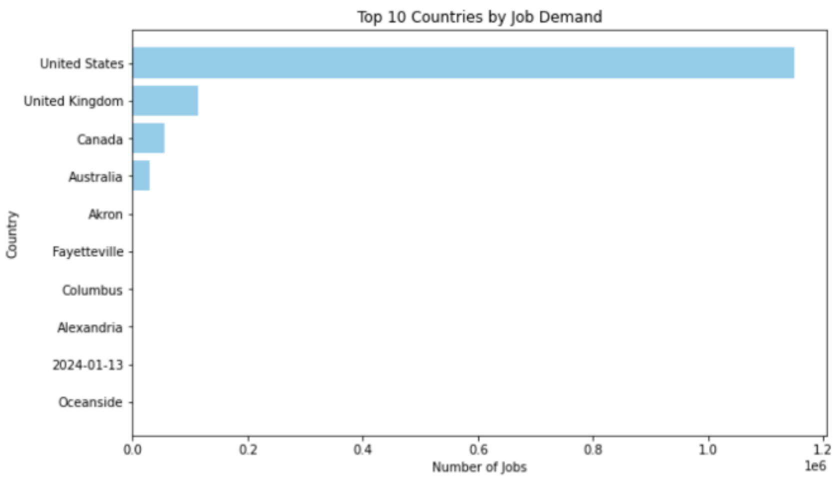

Fig. 3. Top 10 countries by job demand

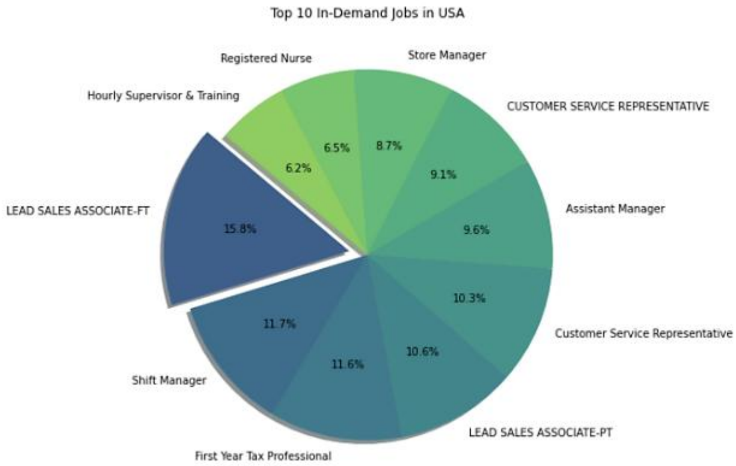

Fig. 4. Top 10 demand jobs in the USA

emphasizing the importance of managerial roles in retail or service sectors. This insight is valuable for workforce planning and job recommendation systems, particularly for targeting high-demand roles.

Figure 5 shows the analysis of the top ten most in-demand skills in job postings. It highlights that Communication stands out as the most prominent. Problem-solving and Customer Service follow closely, showcasing their importance in the workforce. Skills such as Teamwork and Leadership also feature prominently, reflecting their universal value across industries. Additionally, competencies like Time Management, Attention to Detail, and Project Management highlight the need for organizational and strategic abilities. Specialized skills, such as Patient Care, emphasize sector-specific requirements, adding depth to the skill landscape.

In addition, we explored the top ten companies with the most job openings on LinkedIn, led by Health eCareers. Opportunities span diverse sectors, including healthcare, retail, and energy, with notable contributions from Jobs for Humanity and TravelNurseSource. Retail and specialized roles are evident in companies like Dollar General and ClearanceJobs. Food service jobs are highlighted by McDonald's as you see in Figure 6. This distribution reflects active recruitment across various industries. Furthermore, the most required jobs are investigated.

Figure 7 displays the top ten job titles, emphasizing Customer Service Representative as the most prevalent, comprising 25.6%. Lead Sales Associate-FT, Shift Manager, and Store Manager follow as significant contributors, each

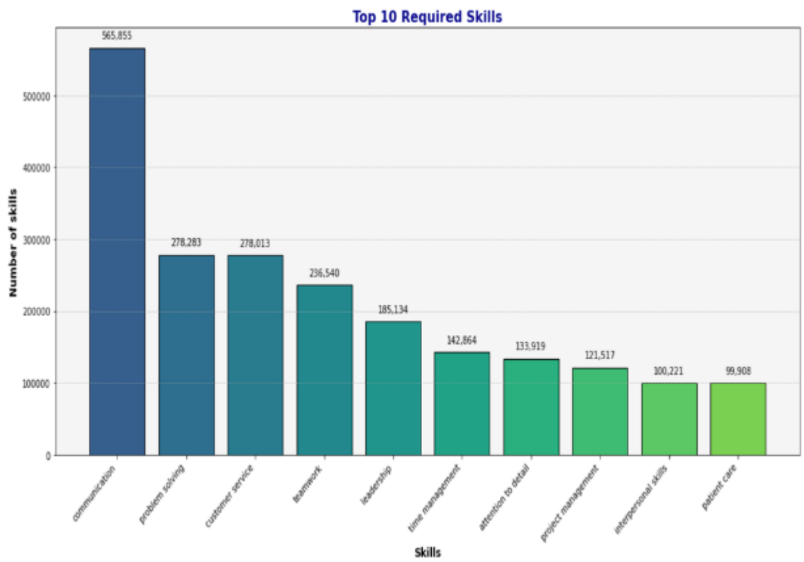

Fig. 5. Top 10 required skills

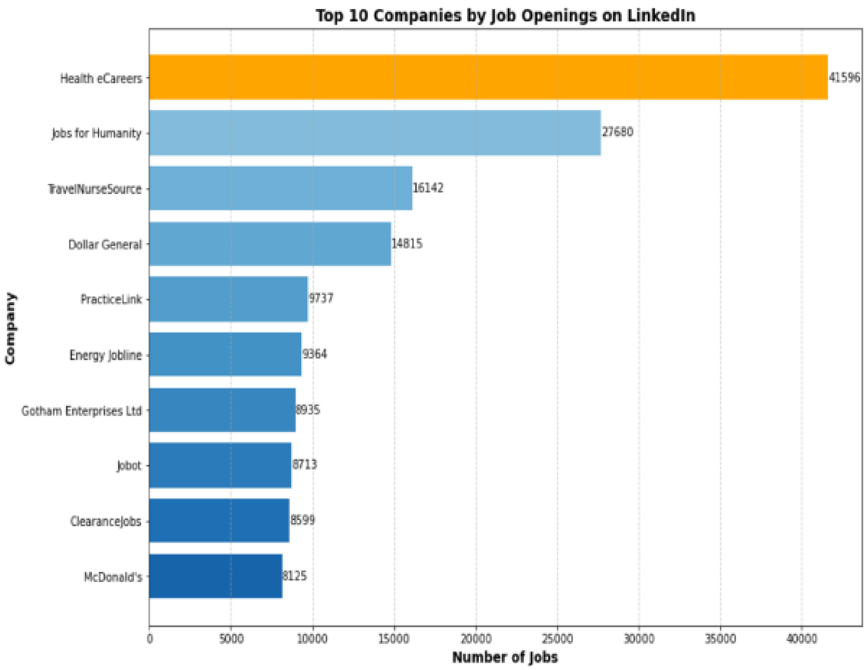

Fig. 6. Top 10 companies by job openings

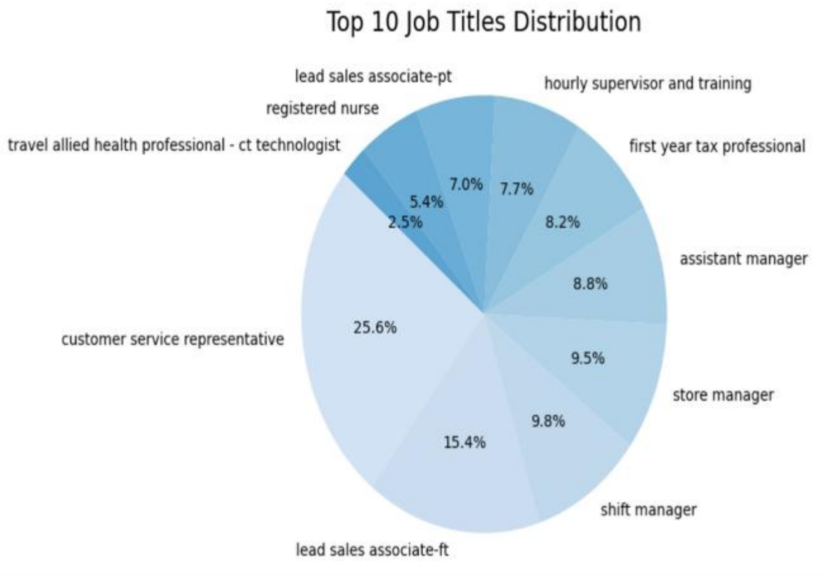

Fig. 7. Top 10 job titles

accounting for 9–15%. Smaller segments include healthcare-related roles such as Registered Nurse and Travel Allied Health Professional - CT Technologist, with shares below 6%. The distribution reflects a mix of customer service, retail, and healthcare roles.

THE PROPOSED SYSTEM

In this paper, a BERT-based Job Recommendation System was built using Apache Spark; Figure 8 shows the general system architecture. The LinkedIn Job Dataset is the source of job postings used in the recommendation process. Data Preprocessing is the first step in the proposed recommender system, with the main role being to combine the job title and job skills of each job posting into job text as explained in the data preparation section. After that, the BERT Embeddings of Job Skills component converts the job text into embeddings using the Sentence-BERT model (all-MiniLM-L6-v2), which transforms textual data into high-dimensional embeddings; these embeddings capture the semantic meaning of the text.

To generate recommendations for a specific user, the user skills are preprocessed and transformed into BERT embeddings, which are dense vector representations. This transformation is performed for effective similarity comparison of user skills against job requirements. After that, Cosine Similarity Calculating is applied. This step involves calculating the cosine similarity between the user's skill embeddings and the embeddings of job text (job title and job skills). Cosine similarity measures the angle between these two vectors, indicating how closely they align to get the most relevant job posting for the user. After calculating similarities,

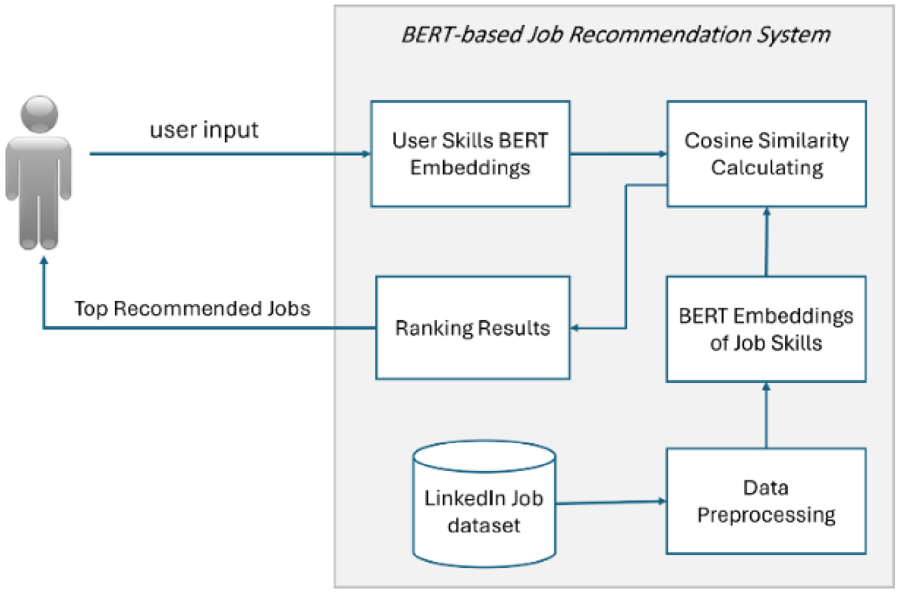

Fig. 8. The proposed system architecture

job postings are ranked based on their alignment with the user's skills. Jobs with the highest similarity scores are prioritized and suggested to the user after being filtered based on user preferences (country, city, job level, job type).

RESULTS AND EVALUATION

The proposed system provides tailored job recommendations by comparing the skills of a user with the requirements of job postings. The BERT model's ability to capture nuanced relationships ensures that synonyms or closely related terms such as "Data Science" vs. "Data Analysis" are treated appropriately, enhancing the matching process. The BERT model was employed to calculate a semantic similarity score between the user's skills and the skills required by each job posting. The similarity score is the backbone of the recommendation process, as it determines the ranking of jobs for the user. A similarity score, such as 0.847 for "Lead Data Scientist," indicates a high level of alignment between the user's skills and the job requirements.

The recommender system provides a range of recommendations tailored for Users. The proposed model presents the ten best-matched job recommendations for different users. These recommendations are tailored to each user by aligning their skills and preferences with the requirements and characteristics of the available job postings. For example, the top recommendation for User ID-2 was Data Scientist (See Table 1). This role appears multiple times with varying requirements, reflecting its alignment with the user's skills.

User skills include "Big Data," "Data Science," "Machine Learning," and "Data Analysis," which indicate a strong technical foundation, particularly in areas relevant to data engineering and analysis. The job skills overlap in "Data Science," "Machine Learning," and "Statistics" which drive higher similarity scores. The Highest Similarity Score was 0.887 for Data Scientist, Arlington. Lower scores for specialized positions such as "Bioinformatics Analyst" suggest the user may benefit from domain-specific upskilling, such as bioinformatics or cloud computing.

Table 1 highlights the best-fitting job for ten individuals, determined by the degree of similarity calculated using the PERT model. This similarity score reflects how well each job aligns with the individuals' skills and preferences, ensuring personalized and accurate recommendations.

Displaying the top recommendations based on user preferences creates a more tailored and effective job search experience. By focusing on relevant criteria such as country, city, job level, and job type, these recommendations enhance user satisfaction, engagement, and overall success in finding suitable employment opportunities. Figure 9 shows the output of the top 5 recommendations for user 1 and user 8 with their different preferences.

For the evaluation process, Precision, Recall, and NDCG (Normalized Discounted Cumulative Gain) metrics were employed to evaluate the proposed system performance. These metrics are common in assessing the performance of recommendation systems [26]. (1) Precision measures the proportion of recommended items that are relevant. In the context of a job recommender system, it indicates what percentage of the recommended jobs are relevant to the user. Further, (2) Recall is a measure that concerns the proportion of relevant items that were recommended. It indicates how many of the relevant jobs were captured by the recommender system. Lastly, (3) NDCG is used to evaluate the quality of ranking in recommender systems. It rewards the model for placing relevant items higher in the recommendation list. The higher the position of a relevant job, the greater the gain.

The evaluation of the proposed model was conducted across three random datasets, each containing 50,000 users. The average for each metric was then calculated. The results of the evaluation were as follows. With an average Precision of 52.67%, the model delivered over half of its top 5 recommendations as relevant, but this drops to 36.67% for the top 10, reflecting a decrease in relevance as the list expands. Recall metrics were notably strong, with an average of 65.33% of relevant items captured in the top 5 and 76.67% in the top 10, indicating the model's ability to cover a broad range of relevant options. Moreover, consistent NDCG scores of 76.67% for both the top 5 and top 10 highlight the model’s strength in prioritizing relevant items at the top of the recommendation list, ensuring users find high-value suggestions quickly. These results suggest that while there is room for improvement in precision, the model effectively balances relevance and comprehensiveness, offering valuable recommendations to users.

TABLE I. JOB RECOMMENDATIONS FOR 10 USERS

User id	Job title	Company	Similarity
1	Lead Data Scientist	Vistra Corp.	0.847
2	Data Scientist	Futran Solutions	0.887
3	Gestionnaire, Communications	Canadian Red Cross	0.780
4	Team Manager	Dunham's Sports	0.819
5	Assistant Manager	Regal	0.738
6	Account Executive	PROS	0.827
7	Finance Manager	Crisis24	0.827
8	Account Executive	Gallin Associates	0.819
9	Registered Nurse/RN	RCM Health Care Services	0.859
10	Attorney	Jobot	0.859

```
User 1:
sn: 1
search_city: Corinth
search_country: United States
job_type: Onsite
job_level: Mid senior
Filtered Recommendations:
+-----------------------------+----------+
|job_title                    |similarity|
+-----------------------------+----------+
|Lead Data Scientist          |0.8469609 |
|Finance Manager              |0.70424205|
|TDM Engineer                 |0.6463835 |
|Application Development Analyst|0.6351797 |
|Sr Software Engineer         |0.6037104 |
+-----------------------------+----------+

User 8:
sn: 8
search_country: United States
job_type: Hybrid
Filtered Recommendations:
+-----------------------------+----------+
|job_title                    |similarity|
+-----------------------------+----------+
|Account Executive            |0.8194624 |
|Account Executive            |0.7628677 |
|Major Account Sales Executive|0.75472254|
|Account Executive            |0.73534584|
|Territory Account Executive  |0.73161966|
+-----------------------------+----------+
```

Fig. 9. Top 5 recommendations with user preferences

CONCLUSION

In conclusion, this paper demonstrated how a BERT-based job recommendation system can effectively match job seekers with relevant opportunities by understanding the context of both job descriptions and user profiles, enabling more precise and context-aware matching between job seekers and available opportunities, and offering relevant and personalized recommendations to users. The evaluation results of the proposed model show that it provides accurate and meaningful matches, with good performance in precision, recall, and NDCG metrics. The model provided relevant recommendations for the top results, but relevance decreased as the list expanded. This work highlights the potential of advanced NLP techniques to improve job recommendations on platforms like LinkedIn, making the job search process more efficient and tailored to individual needs in today's digital workforce. Future work could focus on integrating hybrid recommendation techniques, such as combining collaborative filtering with content-based methods, to improve recommendations for new users or jobs with data scarcity.

REFERENCES

[1] Mueller, A. I., & Spinnewijn, J. (2023). Expectations data, labor market, and job search. In Handbook of Economic Expectations (pp. 677–713). Elsevier.

[2] Siting, Z., Wenxing, H., Ning, Z., & Fan, Y. (2012, July). Job recommender systems: A survey. In 2012 7th International Conference on Computer Science & Education (ICCSE) (pp. 920–924). IEEE.

[3] D. Ganapathy and S. Deepak, "A study on reskilling and networking on Linkedin on employee recruitment success and career advancement," in International Conference on Economics, Business and Sustainability, Springer, 2023, pp. 248–256.

[4] Bianchi, M., Cesaro, F., Ciceri, F., Dagrada, M., Gasparin, A., Grattarola, D., ... & Cella, L. (2017). Content-based approaches for cold-start job recommendations. In Proceedings of the Recommender Systems Challenge 2017 (pp. 1-5).

[5] Yadalam, T. V., Gowda, V. M., Kumar, V. S., Girish, D., & Namratha, M. (2020, June). Career recommendation systems using content-based filtering. In 2020 5th International Conference on Communication and Electronics Systems (ICCES) (pp. 660–665). IEEE.

[6] Sinha, B. B., & Dhanalakshmi, R. (2019). Evolution of recommender system over the time. Soft Computing, 23(23), 12169-12188.

[7] Kanwal, S., Nawaz, S., Malik, M. K., & Nawaz, Z. (2021). A review of text-based recommendation systems. IEEE Access, 9, 31638-31661.

[8] Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[9] T. Erl, W. Khattak, and P. Buhler, Big data fundamentals: concepts, drivers & techniques. 2016. [Online]. Available:

https://sdl.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edshlc&AN=edshlc.014595376.9&site=eds-live

[10] S. Sagiroglu and D. Sinanc, “Big data: A review,” in 2013 international conference on collaboration technologies and systems (CTS), IEEE, 2013, pp. 42–47.

[11] [11] Goundar, S., Bhardwaj, A., Singh, S., Singh, M., & Gururaj, H. L. (2021). Big data and big data analytics: A review

[12] of tools and its application. Applications of Big Data in Large-and Small-Scale Systems, 1-19.

[13] C. A. Udeh, O. H. Orieno, O. D. Daraojimba, N. L. Ndubuisi, and O. I. Oriekhoe, “Big data analytics: a review of its transformative role in modern business intelligence,” Computer Science & IT Research Journal, vol. 5, no. 1, pp. 219–236, 2024.

[14] A. O. Adewusi, U. I. Okoli, E. Adaga, T. Olorunsogo, O. F. Asuzu, and D. O. Daraojimba, “Business intelligence in the era of big data: a review of analytical tools and competitive advantage,” Computer Science & IT Research Journal, vol. 5, no. 2, pp. 415–431, 2024.

[15] Roy, D., Srivastava, R., Jat, M., & Karaca, M. S. (2022). A complete overview of analytics techniques: descriptive,

[16] predictive, and prescriptive. Decision intelligence analytics and the implementation of strategic business

[17] management, 15-30.

[18] Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), Recommender systems handbook (pp. 73–105). Springer.

[19] Singh, A., Rose, C., Visweswariah, K., Chenthamarakshan, V., & Kambhatla, N. (2010, October). PROSPECT: a system for screening candidates for recruitment. In Proceedings of the 19th ACM international conference on Information and knowledge management (pp. 659-668).

[20] Almalis, N. D., Tsihrintzis, G. A., Karagiannis, N., & Strati, A. D. (2015, July). FoDRA—A new content-based job recommendation algorithm for job seeking and recruiting. In 2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA) (pp. 1-7). IEEE.

[21] Nadar, T. K., Athulya, S. S., Rahul, S. S., & Meenakshi, K. (2024, October). Job recommendation by content filtering using TF-IDF and cosine similarity. In 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) (pp. 1585–1590). IEEE.

[22] Panchasara, S., Gupta, R. K., & Sharma, A. (2023, August). AI Based Job Recommedation System using BERT. In 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA) (pp. 1-6). IEEE.

[23] Valgimigli, L. (2019). Job recommendation based on deep learning methods for natural language processing (Master’s thesis, University of Bologna, Campus of Cesena).

[24] Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. Applied Artificial Intelligence, 17(5–6), 375–381.

[25] Asaniczka, “1.3M Linkedin Jobs & Skills (2024),” 2024. doi: 10.34740/kaggle/ds/4418374.

[26] Jadon, A., & Patil, A. (2024, January). A comprehensive survey of evaluation techniques for recommendation systems. In International Conference on Computation of Artificial Intelligence & Machine Learning (pp. 281-304). Cham: Springer Nature Switzerland.