King Saud University
College of Computer and Information Sciences
Department of Information Technology
IT 362 – Data Science
2nd Semester 1447 H

Tech Job Skills Analysis in Saudi Arabia

Section #66849 Group#6	Student Name	Student ID
	Sara Aljuraybah	445201345
	Haya Alomar	445203016
	Fatimah Alhebshi	445200431
	Dalal Alghumlas	445200220

Supervised by: Dr. Reem Alqifari

Contents

1. Introduction

The job market is undergoing continuous change, particularly within technology-driven sectors, where evolving digital and analytical skill requirements make it increasingly difficult for job seekers, educators, and policymakers to identify which competencies are currently most in demand. In Saudi Arabia, understanding skill demand in technology-related roles is especially important for aligning workforce capabilities with the needs of a rapidly growing digital economy. Online job postings for data, artificial intelligence, and related technology positions provide a rich and up-to-date source of information on employer skill requirements; however, this information is not always systematically analyzed to extract actionable insights. Accordingly, this project seeks to address this gap by answering the following research question: What skills are most in demand in technology-focused job roles in the Saudi job market based on online job postings?

2. Literature Review
2.1 Skill Categorization Methods in Job Advertisements [1]

⑩ Problem

The study addressed the problem of identifying the most effective method for extracting and categorizing skill requirements from online job advertisements. Given the variety of existing approaches, the authors aimed to evaluate which method best captures economically meaningful skill information by measuring its explanatory power in wage regressions.

⑩ Dataset

The study used a large-scale dataset of over one million online job advertisements from the UK labor market. The dataset included job titles, detailed job descriptions, and wage information. These advertisements provided unstructured textual data from which skill requirements were extracted and later analyzed in relation to wage variation.

⑩ Methods

The study compared two main approaches for skill extraction: a top-down method based on predefined skill dictionaries and a bottom-up approach using Latent Dirichlet Allocation (LDA) topic modeling. The extracted skills were incorporated into wage regression models to evaluate their explanatory power. The effectiveness of each method was assessed based on its ability to explain wage variation.

⑩ Findings

The study found that the bottom-up LDA-based method outperformed the top-down dictionary approach in explaining wage variation. The LDA method demonstrated higher explanatory power in wage regression models, suggesting that unsupervised skill extraction techniques capture economically meaningful skill patterns more effectively. The results also confirmed that online job advertisements serve as a reliable source for analyzing labor market skill requirements.

2.2 An investigation of skill requirements for business and data analytics positions: A content analysis of job advertisements [2]

Ⓞ Problem

The study investigates the skill requirements for business and data analytics job positions in the United States. Specifically, it aims to clarify the differences and similarities in required skill sets across four analytics-related job categories: Business Analyst (BA), Business Intelligence Analyst (BIA), Data Analyst (DA), and Data Scientist (DS). The study addresses the lack of clear classification and structured mapping of skills associated with these roles in the analytics domain

Ⓞ Dataset

The dataset consisted of 1,235 online job postings collected from Indeed.com over a three-month period (December 2016 to February 2017). The study focused on job postings from four U.S. states: Arkansas, Florida, Kansas, and Missouri. Web scraping and the Indeed API were used to collect full job descriptions for content analysis

Ⓞ Methods

The authors applied content analysis as the primary method. They developed a classification framework consisting of multiple skill categories (e.g., decision-making, programming, statistics, structured data management, communication, organization).

Using text processing techniques, they extracted unigrams, bigrams, and trigrams from job descriptions and matched them against predefined skill keywords. The study then calculated the frequency of skills and skill categories across job postings. Additionally, pairwise comparisons were conducted between BA and BIA, and between DA and DS.

Ⓞ Findings

The findings indicate that decision-making skills are the most demanded across all four job categories. However, differences were observed in technical skill requirements:

- Data Scientist positions require stronger programming and statistical skills.

- Data Analyst roles emphasize structured data management and statistics.
- Business Analyst roles focus more on domain knowledge and soft skills.
- Business Intelligence Analyst roles require a combination of structured data management and statistical tools.

> The study concludes that Data Scientist positions are the most technically demanding, while Business Analyst roles rely more heavily on business domain knowledge and communication skills.

2.3 Demand for Skills on the Labor Market in the IT Sector [3]

⑩ Problem

The study addressed the challenge of identifying key skillsets demanded by employers in the IT sector. Given the rapid evolution of technical and soft skill requirements, the authors aimed to develop a systematic method for extracting and standardizing skill information from unstructured online job advertisements. The study specifically sought to determine which skills and combinations of skills are most demanded across different IT occupational groups.

⑩ Dataset

The study used a large dataset of online job vacancies obtained from the HeadHunter API, the largest recruitment platform in Russia. The dataset focused on IT-related job postings in Saint Petersburg between 2015 and 2019. After filtering and regrouping, approximately 56,000 IT vacancies were analyzed. The dataset included structured information (e.g., vacancy IDs and specialization codes) and unstructured text fields such as job descriptions and listed key skills.

⑩ Methods

The study applied Natural Language Processing (NLP) techniques to extract and standardize skills from unstructured job advertisements. The methods included TF-IDF, tokenization, and the construction of uni-grams, bi-grams, and tri-grams. Skills were standardized using manual synonym matching and similarity measures. Jaccard similarity was used to identify and match related skill terms and to extract significant pairs and triplets of skills across occupational groups.

⑩ Findings

The study identified 435 standardized skills extracted from unstructured IT job advertisements and revealed the most frequently demanded technical and non-technical skills in the sector, such as HTML/CSS, JavaScript, SQL, and Python. Beyond individual skills, the study successfully extracted occupation-specific skill combinations (pairs and triplets) that differentiate between IT occupational groups. The results demonstrated that certain skillsets strongly characterize specific job categories, and that combining similarity measures with n-gram analysis enables effective standardization and classification of skills. The findings confirm that online job advertisements provide a reliable source for identifying both core and occupation-specific skill demand in the IT labor market.

2.4 Organizational and End-User Information Systems Job Market: An Analysis of Job Types and Skill Requirements [4]

⑩ Problem
The study addresses the lack of systematic understanding regarding the structure of the Organizational and End-User Information Systems (OEIS) job market. While substantial research had examined traditional technical Information Systems (IS) roles, there was limited empirical evidence identifying the specific job categories, skill requirements, and responsibility profiles associated with OEIS positions. The authors sought to determine how the OEIS job market is structured, what types of positions exist within it, and what competencies employers explicitly demand in job advertisements. The study aimed to provide clarity for curriculum design, career planning, and alignment between academic programs and labor market expectations.

⑩ Dataset
The dataset consisted of 484 job advertisements collected from U.S.-based online employment sources. These advertisements specifically targeted Organizational and End-User Information Systems roles rather than purely technical programming or software engineering positions. Each advertisement was treated as a unit of analysis. The dataset included information about job titles, responsibilities, required skills, educational background, work experience expectations, and company characteristics

⑩ Methods

The study employed a content analysis methodology to systematically examine the job advertisements. The authors developed a classification framework to categorize positions into job types and subcategories. Through iterative coding and refinement, job titles and descriptions were analyzed to identify recurring patterns in responsibilities and skill requirements. Positions were grouped into broader categories and further subdivided into more specific job classifications. Quantitative frequency analysis was then applied to determine the relative demand for each job type and associated skills. The analysis also examined experience requirements and organizational characteristics to provide contextual understanding of market demand.

⑩ Findings

The study identified five main OEIS job categories, End-User Support, Business Analyst, Training, Web and Interface Design, and Technical Writing, divided into 24 subcategories. End-User Support and Business Analyst roles were most common. Employers emphasized technical, communication, analytical, and business skills, and many positions required prior experience. Overall, the findings highlight a diverse, non-programming-focused job market requiring interdisciplinary competencies.

2.5 Literature Comparison

Across the reviewed studies, a clear commonality is the use of online job advertisements as the primary data source for analyzing labor market skill demand. All four studies rely on unstructured textual data from job descriptions and titles, demonstrating the value of digital job platforms for extracting employer-required skills. Despite differences in geographic focus (UK, United States, and Russia) the studies consistently confirm that online job postings provide rich and scalable data for labor market analysis.

However, the studies differ significantly in their analytical focus and modeling approaches. Study [1] evaluates competing skill extraction methodologies by comparing dictionary-based methods with LDA topic modeling and validating results using wage regression. Study [2] applies structured content analysis and predefined skill frameworks to compare analytics-related job roles. Study [3] employs NLP techniques such as TF-IDF, n-grams, and similarity measures to standardize and cluster IT skills and identify occupation-specific skill combinations. In contrast, Study [4] focuses primarily on developing a classification framework to structure the OEIS job market, using iterative content analysis and frequency analysis rather than advanced computational modeling. This reflects a shift from methodological optimization and skill extraction toward market structure and job taxonomy development.

While each study contributes valuable insights, several limitations remain. Most studies are geographically limited and may not generalize across labor markets. Some approaches rely on predefined skill taxonomies, potentially restricting the identification of emerging skills. Additionally, none of the reviewed studies examine the Saudi Arabian labor market, particularly within technology and data-related roles.

Overall, although prior research demonstrates the effectiveness of using online job advertisements to analyze skill demand, there remains a gap in localized, technology-focused analysis within Saudi Arabia. Addressing this gap can provide clearer insight into employer requirements in the Saudi data and AI job market.

3. Data Source:

Job postings data was collected using the JSearch API provided via RapidAPI. The dataset consists of technology-related job advertisements published on multiple job platforms (such as LinkedIn and other job boards) and filtered specifically for positions located in Saudi Arabia. The collected data focuses on roles within the data and AI domains, including Data Scientist, Data Analyst, and related positions.

- Data Collection:
  The data was collected by sending HTTP GET requests to the JSearch API endpoints using Python. Specific query parameters were used to retrieve job postings related to data and AI roles within Saudi Arabia. The API responses were returned in JSON format and then converted into a structured tabular format for further analysis. Multiple pages of results were fetched to increase the number of job postings collected.

```python
response = requests.get(url, headers=headers, params=params)
data = response.json()
```

3.1 Data Description

1. Number of Observations: 1025
2. Number of features: 31
3. Data Types:

Feature	Data Type	Description
job_id	Qualitative (Nominal)	Unique identifier for each job posting.
job-title	Qualitative (Nominal)	Title of the advertised job position.
employer_name	Qualitative (Nominal)	The name of the company offering the job.
employer_logo	Qualitative (Nominal)	Link to the employer's logo image.
employer_website	Qualitative (Nominal)	Official website of the employer.
job_publisher	Qualitative (Nominal)	Platform where the job was published.
job_employment_type	Qualitative (Nominal)	Primary employment type of the job.
job_employment_types	Qualitative (Nominal)	All employment types associated with the job.

job_apply_link	Qualitative (Nominal)	URL used to apply for the job.
job_apply_is_direct	Qualitative (Nominal)	Indicates whether the application is direct.
job_city	Qualitative (Nominal)	City where the job is located.
job_country	Qualitative (Nominal)	Country where the job is located.
job_latitude	Quantitative (Ratio)	Latitude of the job location.
job_longitude	Quantitative (Ratio)	Longitude of the job location.
job_description	Qualitative (Nominal)	Textual description of the job role.
job_benefits	Qualitative (Nominal)	Benefits offered with the job.
job_google_link	Qualitative (Nominal)	Google Jobs link for the posting.
job_salary	Quantitative (Ratio)	Text-based salary information.
job_min_salary	Quantitative (Ratio)	Minimum salary offered.
job_max_salary	Quantitative (Ratio)	Maximum salary offered.

job_salary_period	Qualitative (Nominal)	Salary payment period.
job_onet_soc	Qualitative (Nominal)	O*NET occupational classification code.
job_onet_job_zone	Quantitative (Ordinal)	Experience and preparation level indicator.
search_query	Qualitative (Nominal)	Search keyword used to retrieve the job.
job_posted_at_datetime_utc	Quantitative (Interval)	Job posting date and time (UTC).
job_posted_at_timestamp	Quantitative (Ratio)	Timestamp of job posting.
job_offer_expiration_datetime_utc	Quantitative (Interval)	Job offer expiration date (UTC).
job_offer_expiration_timestamp	Quantitative (Ratio)	Timestamp of job offer expiration.
job_required_experience	Qualitative (Ordinal)	Required experience level.
job_required_skills	Qualitative (Nominal)	Required or preferred skills.
job_required_education	Qualitative (Ordinal)	Required education level.

3.2 Dataset Bias Evaluation:

- **Representation Bias:**
  The dataset is collected from online job platforms through the JSearch API, which may overrepresent technology roles from large companies and urban areas while underrepresenting smaller organizations or roles advertised through informal or offline channels.

- **Measurement Bias:**
  Some variables, such as salary information and required skills, are inconsistently reported across job postings. In addition, skill requirements are extracted from textual descriptions, which may vary in detail between employers.

- **Historical Bias:**
  The dataset may reflect existing inequalities in the job market, such as greater visibility of roles requiring English proficiency or advanced technical backgrounds, potentially limiting representation of certain groups or regions.

4. Objectives

- To identify the most frequently demanded skills in data and AI–related job postings in Saudi Arabia.
- To examine the distribution of different skill categories, such as technical, analytical, and soft skills, within Saudi data and AI job advertisements.
- To analyze variations in skill requirements across different data and AI roles, including positions such as Data Scientist and Data Analyst.
- To explore common combinations of skills requested by employers in data and AI job postings in the Saudi job market.

5. Methods

1. Data Preprocessing

First, duplicate entries were identified and removed to prevent inflation of skill frequency counts and role distribution analysis. Duplicates existed both at the record identifier level and at the content level (repeated postings with identical core information). Retaining only unique postings ensured analytical validity. Second, missing values were evaluated at the attribute level. Fully empty attributes were excluded from analysis, while partially missing information was retained when relevant to the study objectives to preserve dataset size and representativeness.

Third, textual cleaning was applied to remove non-informative artifacts introduced during web scraping and to standardize text formatting. Consistent casing and normalization were applied to ensure uniform representation of English skill terms.

Finally, terminology normalization was performed to address variations in how the same skills were written (e.g., full terms and abbreviations). Equivalent expressions were unified under standardized labels before frequency computation to avoid fragmented counts and inaccurate demand estimation.

2. Data Transformation:

The transformation process focuses on organizing the dataset into structured and analyzable formats aligned with the research objectives.

Skills will be standardized to unify different expressions of the same concept (e.g., abbreviations and full terms) to ensure accurate frequency measurement and pattern detection.

Job roles will be grouped into consistent categories to enable comparison of skill requirements across different data and AI positions.

Skill categories will classify extracted skills into broader groups such as technical, analytical, and soft skills to support proportional distribution analysis.

Finally, the structured dataset will be stored in formats such as CSV or JSON to ensure compatibility with analytical and data mining tools.

3. Analysis Techniques:

   Once the data is preprocessed and transformed, a set of analytical techniques will be applied to achieve the project objectives related to skill demand in data and AI job postings in Saudi Arabia.

   Skill frequency analysis will be used to extract skills from job descriptions and required skills fields and count their occurrences across job postings. This technique directly supports identifying the most frequently demanded skills in the Saudi job market.

   Skill category distribution analysis will be conducted by grouping extracted skills into technical, analytical, and soft skill categories and calculating their relative proportions. This analysis supports understanding the balance of different skill types requested by employers.

   Role-based comparative analysis will analyze differences in skill requirements across data and AI roles by grouping job postings based on job titles such as Data Scientist and Data Analyst. This technique supports examining variations in skill demand across roles.

   Skill co-occurrence analysis will be applied to identify common combinations of skills that appear together within the same job postings. This analysis supports exploring how employers bundle skills when advertising data and AI positions.

4. Reporting insights:

   The analysis results will be reported in a clear and structured manner to address the project objectives related to skill demand in data and AI job postings in Saudi Arabia.

   Frequently demanded skills will be presented using ranked summaries to highlight the most in-demand competencies. The distribution of skill categories will be summarized to show the balance between technical, analytical, and soft skills. Variations in skill requirements across different data and AI roles will be reported through comparative summaries. In addition, common combinations of skills requested by employers will be highlighted to reflect typical skill sets in job postings.

   Overall insights will be summarized in relation to the Saudi job market context.

6. Challenges Faced and Recommendations

During the data collection phase, several challenges were encountered when using the JSearch API. The API subscription was limited, which restricted the number of requests and the amount of data that could be collected. In addition, the data retrieval process was time-consuming due to API response time and pagination when fetching multiple job postings. To improve future work, it is recommended to use a higher-tier API plan, extend the data collection period, and integrate additional job platforms to obtain a larger and more representative dataset.

In addition, future efforts should consider expanding data sources to reduce platform-specific bias and improve market representativeness. Extending the data collection period would also help minimize temporal bias and provide a more stable view of skill demand. Enhancing cross-platform deduplication and refining skill extraction methods would further strengthen data accuracy and analytical reliability.

7. References

[1] Z. Ao, G. Horvath, C. Sheng, Y. Song, and Y. Sun,
"Skill requirements in job advertisements: A comparison of skill-categorization methods based on explanatory power in wage regressions,"
arXiv preprint arXiv:2207.12834, 2022.
Available: https://arxiv.org/abs/2207.12834

[2] S. Verma, A. Chandra, and R. Kumar,
"An investigation of skill requirements for business and data analytics positions: A content analysis of job advertisements,"
Journal of Education for Business, vol. 94, no. 6, pp. 345–357, 2019.
Available: Reserch paper link

[3] Ternikov A. A. and Aleksandrova E. A., "Demand for skills on the labor market in the IT sector", *JBI*, vol. 14, no. 2, pp. 64-83, Jun. 2020.
Available: https://doi.org/10.17323/2587-814X.2020.2.64.83

[4] C.–A. Chao and S. C. Shih,
"Organizational and End-User Information Systems Job Market: An Analysis of Job Types and Skill Requirements,"
Information Technology, Learning, and Performance Journal, vol. 23, no. 2, pp. 1–15, 2005.
Available:
https://www.researchgate.net/publication/242144668_Organizational_and_End-User_Information_Systems_Job_Market_An_Analysis_of_Job_Types_and_Skill_Requirements