



USING DATA MINING TO IDENTIFY SKILLS GAPS AND PROVIDE COURSE RECOMMENDATIONS

Diksha Tiwari, Karthikeyan Balakrishnan, Claire Shuo Jin, Matthew A. Lanham
Purdue University, Krannert School of Management
tiwarid@purdue.edu; balakri7@purdue.edu; jin417@purdue.edu; lanhamm@purdue.edu

ABSTRACT

We have developed an AI course consulting tool that helps organizations understand the current skill gap using open job positions and helps them train their workforce accordingly. This research project helps one of the biggest tech workforce development organization gain strategic insights into the most essential skills needed by its clients and provide relevant course recommendations. In this study we explore 1) Techniques to ensure continued scraping and making the scraping process streamlined and production ready. 2) Data mine the job posting data to capture insights and any industry-level trends 3) Develop a text Analytics unit that processes the data and captures insights from it. 4) Summarize the job-board insights using a customer-facing dashboard.

INTRODUCTION

There is an increasing gap developed in which employers struggle to hire appropriately trained workers. According to a McKinsey Global Survey, conducted in Feb. 2020, on future workforce needs, nearly nine in ten executives and managers say their organizations either face skill gaps already or expect gaps to develop within the next five years.

Respondents expect to see skill gaps as market and technology trends alter organizations' talent needs.

When skill gaps are expected to occur within organizations, % of respondents'



Share of organizations' current roles at risk of being disrupted by market or technology trends in next 5 years, % of respondents'



*Figures may not sum to 100%, because of rounding; n = 1,216.

McKinsey & Company

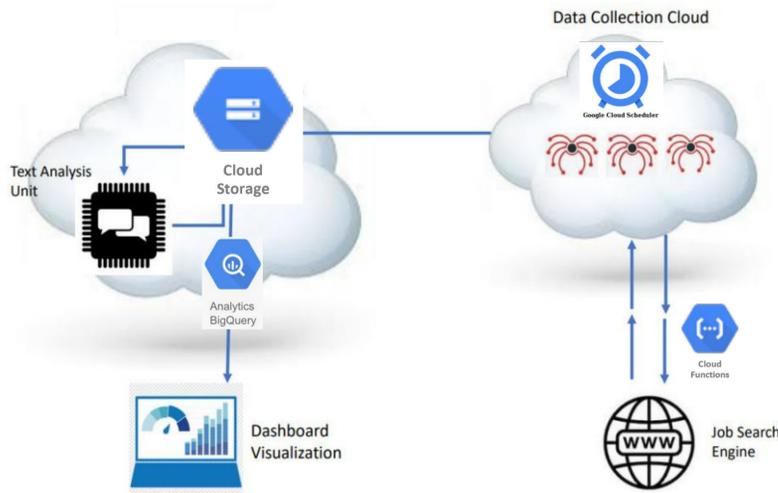
Research Questions:

- How to automate the scraping process to create a data lake for job-postings?
- Can we develop a methodology to extract tech skills from job-postings?
- Can develop a tool that uses tech-skills from job-postings to identify the skill-gap?

LITERATURE REVIEW

Study	text analytics	clustering	web scrapping	forecasting	visualization
Boothby Clara [et al.]	✓				
Zhang Wen [et al.]	✓				
Matt J. Kusner [et al.]	✓				
Heidarysafa Mojtaba [et al.]	✓		✓		✓
Florian Beil [et al.]		✓			
Akshi Kumar [et al.]		✓			
Hing Kai Chan [et al.]				✓	
Our Study	✓	✓	✓	✓	✓

METHODOLOGY



Source: [2106.11077] Toward a Knowledge Discovery Framework for Data Science Job Market in the United States (arxiv.org)



Fig 2. Study Design

Web scrapers: Scrapes job posting websites and collect information such as Job titles, Job Id, Job Category, Job Location, Job Description, Date Posted, and Timestamp.



User Agent



Error Exception



Random Delay / Sleep

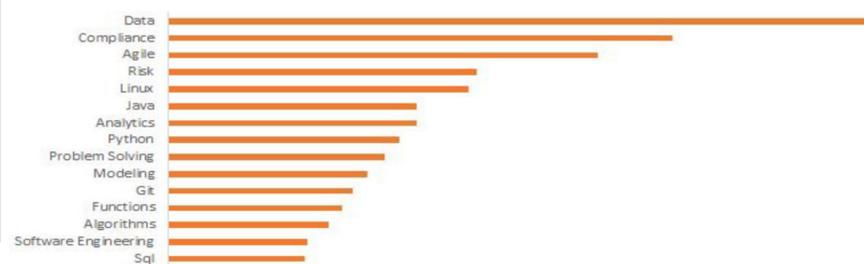
Cloud Function: automates scraping process. The function triggers the scrapers designed to scrape the job postings from the website and store the raw data in the database.



Algorithm – Text Analytics

Data Cleaning: Checks for null and invalid values in the data and cleans it.

Feature engineering: Extracted skills, location, years of experience, seniority and education from job description.



FINAL DASHBOARD

KEY TAKEAWAYS

- Allow client salespeople to identify course opportunities in target clients through data
- Find course opportunities for a client based on their competitor's open job postings
- Help organizations understand the current skill gap using open job positions and train their workforce accordingly.

UNFOCUSED PITCH ❌

"We have over 500 courses! Which ones do you want?"

HIGH-IMPACT PITCH 😊

"Here are the top 10 skills your company is hiring for currently, and here are our select courses we recommend."

ACKNOWLEDGEMENTS

We would like to thank Professor Matthew Lanham, Saurabh Bhise, and Yogesh Tajave for their guidance and support on this project.