Job Market Skills Monitor: An NLP Approach

MSBA 316: Natural Language Processing & Text Analytics

Abbas Nassereddine

9/5/20

To: Dr. Wael Khreich

Abstract

We are truly currently witnessing unprecedented times. The ongoing COVID-19 pandemic that has been sweeping through the world is sending enormous shockwaves throughout a job market that has already been suffering from imbalances in the supply-demand of human resources and available job vacancies. The MENA region has been hit especially hard with the plummeting of oil revenues which have always provided many economies in the region with continuous cash flow, which in turn would create vast employment opportunities in various sectors. Therefore, it has become essential, more than ever, to rethink our job market dynamics. We should be keen to carefully assess job market needs so as to optimize human capital accordingly. This will also give policy makers to quickly detect changes in market skill demand. It is crucial that policy makers adjust university programs and/or enrollment quotas so that job markets needs are satisfied dynamically.

From there, we provide data-driven solutions by leveraging the power of the data already available online. LinkedIn, being one of the most prominent job advertisement platforms, has been chosen as the source of data. We employ state-of-the-art Natural Language Processing algorithms and techniques to collect, process, and dissect information about the state of the job market. Our main metrics of interest are advertised job positions, their respective descriptions as posted by the recruiters, and date of posting. The date of posting of the job provides useful time-series data as to what jobs are demanded during a given time frame. Job descriptions are naturally used to extract information regarding demanded skills and tasks.

In the end, we have managed to aggregate the advertised positions under clean well-defined “umbrellas”, pinpointed in-demand skills for main occupations, presented time-series insights regarding the state of the job market, and used the collection of extracted data to produce recommendations and develop techniques that can be the foundations for future work on the very important topic.

Contents

Table of Figures

Table of Tables

I. Introduction

The MENA region is underperforming in comparison to many other regions in the world when it comes to unlocking the potential of its human capital. According to the World Economic Forum's Human Capital Index, the MENA region only captures 62% of its human potential.

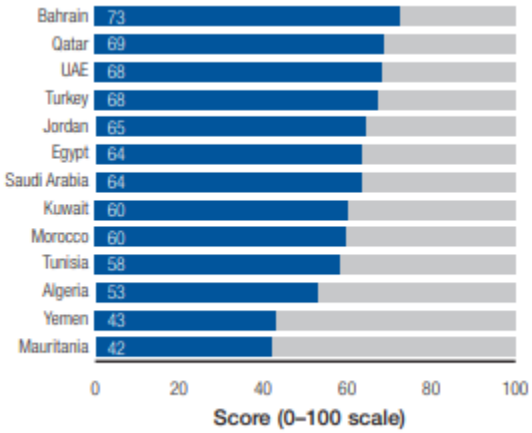


Figure 1 Human Capital Optimization in the MENA region

The World Economic Forum also noted that two in five university graduates are unemployed in the MENA region at any given time (World Economic Forum, 2017). To have counteracted this, the World Economic Forum stresses that at least 5 million people should have been reskilled in the MENA region by 2020.

In many countries in the MENA region, many investments and improvements in the job market had already been underway, but the problem persisted. Therefore, it can be safe to conclude that policy-makers in such countries, and others, are lacking the information necessary to know where and how to direct investments and funds in the job market and education sectors rather than the absence of funds. And unfortunately, the MENA region generally lacks a coherent and solid framework that provides real-time job market trends and metrics.

In another report, titled "Talent Mobility Good Practices", Mercer Mobility, affirms that "87 per cent of CEOs in the Middle East believe that the limited supply of candidates with the right skills will present the biggest organizational challenge over the next three years." (Mercer Mobility, 2012). This shows that senior hiring managers and applicants both benefit from balanced skills supply-job demand.

Saudi Arabia, the region's biggest economy, is in the eye of the storm when it comes to the future of the human resources-job market balance. The country can expect 46% of its work activities to be automated in the near-to-middle future (World Economic Forum, 2017). Therefore, it would be necessary to upskill the workforce in the country to satisfy market demands.

For this reason, this project will aim to lay the foundations and methodologies for a framework that would provide crucial insights regarding the state of the job market at a given time. Also, our framework has the ability to quickly detect a new in-demand skill in the market, so policy makers can quickly act upon this new scenario. The framework will make use of the extensive and dynamic

job advertisements posted online. The extracted data will then be analyzed and dissected using NLP techniques and then finally results will be discussed and recommendation will be given.

II. Literature Review

Previous work by NLP scholars and practitioners has demonstrated the extent to which NLP techniques can present insights on the state of the job market.

Job descriptions contain a variety of job information types ranging from work-related attributes to work-related tasks. In general, employers do not follow a universal format to lists the sought skills and tasks an applicant should do. Therefore, it is necessary to develop a dynamic tool that would allow us to properly segregate whether a given sentence in a job description is a work attribute or a work task. An article titled " Text Mining in Organizational Research" (Kobyashi et al., 2017), discusses an algorithm that would predict whether a given sentence in a job description is a skill or a task using a machine learning model applied on well-defined features extracted from data labelled as work skills and work tasks. The researchers used labelled data from Monsterboard to train the model.

The model uses Part of Speech (POS) distribution and word frequency measures in sentences pertaining to skills versus those pertaining to tasks as features. The power of this model is in that it recognizes the importance of semantics and pragmatics. The researcher was interested in doing further analysis on sentences that pertain to skills rather than tasks which were discarded from the later topic modelling.

The following table displays the features extracted and which were used to train the prediction model:

Table 1 Features used for the Label Classification Algorithm

Feature Type	Number of Derived Features	Variable Type
Part of speech (POS) tag of the first word	1	Categorical (actual POS)
Is the first word in this sentence unique in work activity sentences (based on the labeled data)?	1	Numeric
Is the first word in this sentence unique in worker attribute sentences (based on the labeled data)?	1	Numeric
Is the last word in this sentence unique in work activity sentences (based on the labeled data)?	1	Numeric
Is the last in this sentence unique in worker attribute sentences (based on the labeled data)?	1	Numeric
Proportion of adjectives	1	Numeric
Proportion of verbs	1	Numeric
Proportion of the word "to"	1	Numeric
Proportion of modal verbs	1	Numeric
Proportion of numbers	1	Numeric
Proportion of adverbs	1	Numeric
Proportion of nouns	1	Numeric
Proportion of nouns, verbs, adjectives, adverbs, and other part of speech tags followed by another verb	5	
Proportion of unique words found only in work activity sentences (based on the labeled data)	1	Numeric
Proportion of unique words found only in worker attributes sentences (based on the labeled data)	1	Numeric
Frequency of keywords for work activity and worker attributes sentences	149	Numeric

Three models were used to Support Vector Machine, Random Forest Classification, and Naïve Bayes. Random Forest had proven to be the best performing model of the three. The researchers then applied LDA on the data labelled as "work attributes" (skills) in order to identify clusters of jobs using a k value (number of topics) of 140.

The following sample of topic clusters was identified:

Table 2 Extracted Topic Models From Applying LDA to Worker Skills Sentences

Topic 100	Topic 86	Topic 132	Topic 75
development	new	travel	communication
software	learn	willingness	written
agile	quickly	willing	oral
methodologies	willingness	work	verbal
application	adapt	time	interpersonal
scrum	technologies	needed	presentation
design	internet	internationally	effective
life	desire	international	listening
Topic 18	**Topic 45**	**Topic 20**	**Topic 105**
highly	detail	sales	results
motivated	attention	selling	leadership
oriented	oriented	salesforcecom	others
self	organizational	outside	goals
driven	accuracy	crm	achieve
organized	multitask	success	influence
starter	follow	account	motivate
selfstarter	details	inside	deliver

Another paper delved into the time-series aspect of job market analysis. A faculty member at the department of Computer Science (COSC) at Indiana University of Pennsylvania (IUP), published his findings regarding the IT job market in paper titled "Analyzing Computer Programming Job Trend Using Web Data Mining". The researcher applied Keyword Extraction and frequency counting techniques on job postings posted on a website called Dice.com (Smith et. al, 2014). Smith developed a recurrent process that would send queries to the website and would count the number of results obtained for each query. Then, to normalize the data, the time series collected for each query would be averaged over a month and the monthly frequency behavior would be plotted. The researcher tagged COBOL (business-oriented software), Python, Ruby, and Java, as skills of interest to analyze the number of monthly search results for each skill. The following results were obtained:

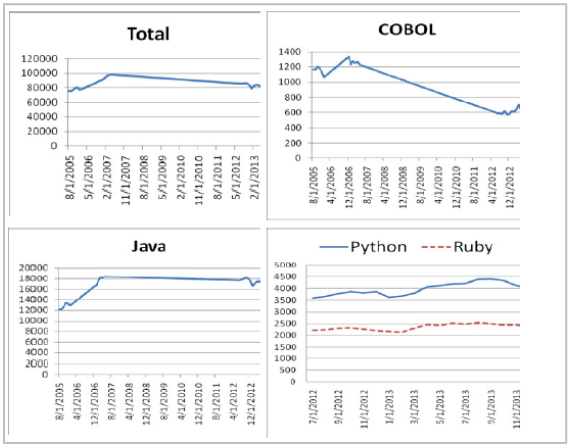


Figure 2 Time Series Describing the Frequency of Instances of COBOL, JAVA, Python, and Ruby

The shortcoming in this strategy is that the researcher is making the assumption that no skill will be mentioned in the postings that he is not familiar with. This compromises the study's ability to capture novel market trends and solely depends on previous (and arguably obsolete) knowledge in skills required by the job market.

In an attempt to further understand the job market dynamics in five US states, a group of researchers published a paper titled "Data Mining Approach to Monitoring The Requirements of the Job Market: A Case Study", where they investigated job market descriptions using Latent Semantic Indexing (LSI). Data about the 1110 recognized occupations in the US, along with their corresponding descriptions were collected from O*NET (a database that contains occupational definitions). To prepare their data, the involved researchers, MacCrocy et al., removed stop words so as to decrease noise, and then removed words that appear only once in their corpus (MacCrocy et al, 2016). Next, the Term Frequency-Inverse Document Frequency (TF-IDF) matrix of the O*NET job descriptions, X, was generated, in order to be later used as input to the LSI model. After that, the matrix was transformed using Singular Value Decomposition (SVD), where X was expressed as

$$X = LSR^T$$

Where:

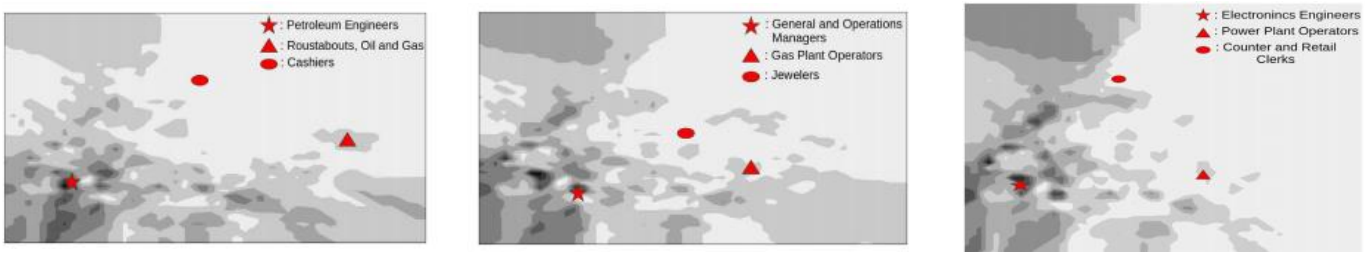
- L is the left singular vector describing the job description-job type relationship
- R is the right singular vector describing the word-job type relationship
- S is the strength vector which measures the importance of each job category in the whole dataset. This is a diagonal matrix and shows the strength values on its diagonal in decreasing order. The model while be optimized by choosing the number of strength factors (and topics) that would minimize error. Other strength factors corresponding to minor job types or noise were omitted.

As a result, the optimum number of general occupational types in the O*NET database was found to be k=69.
The same vector transformation using the same number of optimum k was applied on a list of job descriptions collected from online job advertisement web sites.
Each job posting was projected as follows:

$$Q_k = S^{-1}L^TQ$$

where S-1 is the inverse of the diagonal matrix of singular values S, LT the transposed of the left singular vector's matrix L and Q the term vector

Then, for each O*NET job description, the cosine similarity was found for every Q_k vector obtained from before. Then, after recalculating the set of cosine similarities using the sigmoid function (to emphasize variability of similarities), the similarities are summed over n (number of job postings), and then normalized by dividing by n. The final product is a metric called "Demand Weight", which shows how much a job title defined by O8NET database is prevalent within a dataset.
The entire process was repeated on job postings pertaining to different general work clusters: Oil and Gas , and Banking as well as across different geographical areas: United States, United Kingdom, and GCC region.



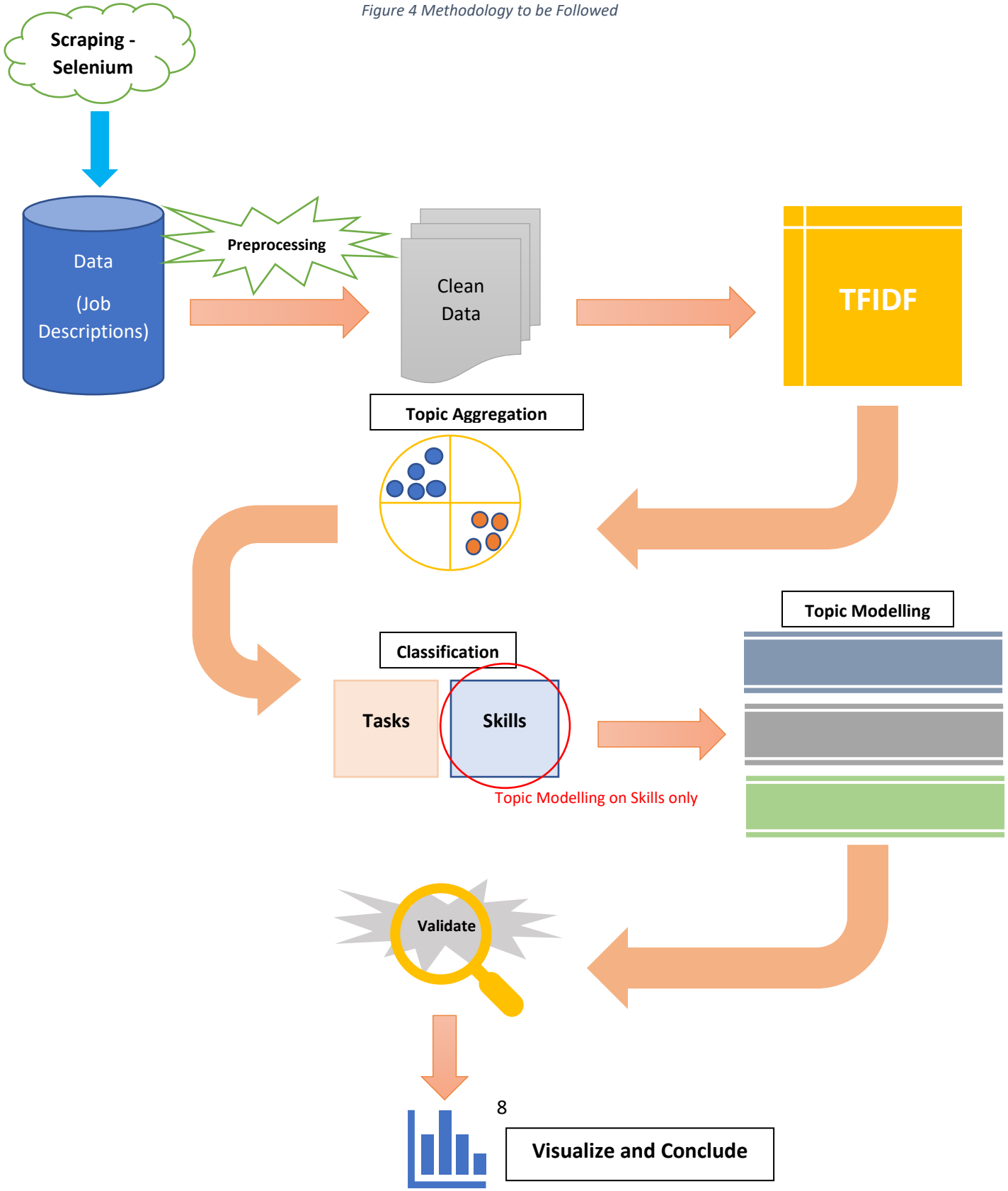
(a) *(b)* *(c)*

Figure 3 The landscape of oil and gas industry jobs in GCC region (a), USA (b), and UK (c)

The above results clearly display the severe oil dependency of the GCC job market on Petroleum Engineers as opposed to Gas Plant Operators in the USA, which is a reflection of each area's power production practices.

III. Methodology

Figure 4 Methodology to be Followed



To complete our analysis, we go through the methodology described in the schematic above. First, we used Selenium to scrape data from LinkedIn, then we clean and pre-processed the collected data according to our requirements. After that, we perform topic aggregation of the many different job titles. Then, we used feature engineering to extract the features that would be used to train a classification model that would distinguish between sentences describing work skills vs. work tasks in a job description. We perform topic modelling per job aggregation on the respective “skills” sentences so as to acquire insights regarding the needed skills for each job position sub-cluster. We also present time-series analysis of job posting trends and of work experience required (using Named Entity Recognition). Finally, we will aggregate all our findings in order to give recommendations regarding the current state of the Saudi job market as extracted from LinkedIn.

A. Data Collection

LinkedIn is the largest platform connecting professionals around the world and one of the most used tools to advertise jobs and for other recruitment processes. Due to this richness of data, LinkedIn can prove to be a valuable source when it comes to providing insight regarding the state of the job market at a given moment. Therefore, many companies may attempt to scrape publicly available data from LinkedIn to provide business intelligence.

One of the most prominent companies scraping LinkedIn data is HiQ. The company, due to its web scraping activities, was involved in a lawsuit with LinkedIn. LinkedIn invocated the Computer Fraud and Abuse Act in a cease-and-desist letter to HiQ (Robertson ,2019). However, HiQ insisted that the data was published by the users under their consent. In the end, the court ruled that since the data is available for anyone to access, and that there has proven to be many benefits for web scraping, that HiQ may continue to scrape the publicly available data from LinkedIn. This court verdict will be leveraged in case any future cases are raised against web scraping.

Therefore, driven by the motivation to innovate and formulate a product that would help policy makers and students, and referring to the above well-renowned court verdict, it is safe to assume that the web scraping we will attempt to do can be justified.

To acquire the necessary data from LinkedIn, a Selenium web crawler was developed. The crawler will sign in using a username and password. And then it will go to the "Jobs" feature in LinkedIn and will input the desired job location, which has been chosen, in our case, to be Saudi Arabia. When the results are loaded, the scraper will extract every job position and its corresponding job description. After each job position extraction, the crawler should also scroll down the page so that the new job positions appear. Also, the crawler will move to the next page once the results on a given page are all extracted.

A total of approximately 1000 rows were extracted each containing details about the advertised Job Position, Job Description, and the time that has passed since this post had been published.

Table 3 Sample of the Data Scraped from LinkedIn

	Job Position	Job Description	Date
50	Demand Planner	Demand PlannerJob Descripti	Posted DatePosted 2 months ag
51	Cementing Field Specialists	We are looking to hire Cemen	Posted DatePosted 2 weeks ago
52	CIB Operations - Securities C	JP Morgan is hiring within the	Posted DatePosted 1 month ago
53	Manager	Manager - 50942Client Service	Posted DatePosted 10 hours ago
54	Internal Communication Spe	Job purpose: The job main rol	Posted DatePosted 1 week ago
55	Link 16 Operation Co-ordinat	Link 16 Operation Co-ordinato	Posted DatePosted 1 week ago
56	Administrative Assistant	Overview / ResponsibilitiesW	Posted DatePosted 2 weeks ago
57	S&OP Program Manager	DescriptionIf you are an exper	Posted DatePosted 9 hours ago
58	Waste Management Speciali	Post: Waste management sale	Posted DatePosted 2 weeks ago
59	APPLICATION DEVELOPPER N	The Application Developer ha	Posted DatePosted 3 months ag
60	Internal Medicine Specialist	Minimum requirements -* Tra	Posted DatePosted 1 day ago
61	Development Manager – Rer	Before you click “Easy Apply”:	Posted DatePosted 1 week ago
62	Executive Program Administ	Job DescriptionAct as the cent	Posted DatePosted 2 months ag
63	Project Coordinator (Jeddah,	SOSi has an immediate need f	Posted DatePosted 1 month ago

As we take a look at the data, it is immediately clear that, naturally, the data is not structured as it has been inputted by human recruiters, each one using varying text structures, job titles, and semantics. For example, different job titles are used to describe jobs that almost totally describe the same position (“Business Finance Analyst” vs “Business Analyst”). This will require tailored preprocessing, aggregation of job titles, and further dissection of job descriptions so that job-specific

B. Pre-Processing

As discussed, dates can provide important insights regarding hiring trend during some time period. However, as extracted from LinkedIn, the dates column contains entries that contain strings such as “months ago”, “days ago”, and “weeks ago”. All the entries had to be converted to numerical values corresponding to days. For example, “7 weeks ago” had to be changed to 28 (7*4).

An important that was also extracted is the required years of experience per job cluster. This data was present inside the job description. Using Named Entity Recognition (NER), we were available to pinpoint entities that pertain to time. However, many of the entries were written in string format (eg.: “four years of experience”). Also, all entries which had a value of above 30 were omitted as they may actually be referencing the date the company advertising the job position was founded, and this is irrelevant.

C. Topic Aggregation

As mentioned, different recruiters have named similar job positions using varying terms. Therefore, it is necessary to acquire representative and well-defined job titles. Differently named but contextually similar job positions will be mapped to have the same tag. To achieve this, a dataset containing branches, sub-branches, and titles of jobs was obtained from O*NET, a database dedicated for providing standard definition and descriptions of job occupations.

A sample of the obtained data is shown here:

Table 4 Dataset obtained that contains Different Levels of Career Clustering

Browse by Career Cluster
All Career Clusters

Career Cluster	Career Pathway	Co	Occupation
Arts, Audio/Video Technology & Communications	Visual Arts	27-1024.00	Graphic Designers
Arts, Audio/Video Technology & Communications	Visual Arts	27-1014.00	Multimedia Artists and Animators
Arts, Audio/Video Technology & Communications	Visual Arts	27-1027.00	Set and Exhibit Designers
Business Management & Administration	Administrative Support	43-3031.00	Bookkeeping, Accounting, and Auditing Clerks
Business Management & Administration	Administrative Support	43-2099.00	Communications Equipment Operators, All Other
Business Management & Administration	Administrative Support	43-9011.00	Computer Operators

The above dataset is a tabulated tree of father nodes of “Career Clusters” and son nodes of “Career Pathway”, which are intern father nodes of “Occupation”. We compared the similarity of each scraped job position with all the formalized job titles. The scraped job title was associated and renamed to the formalized job title with which it shares the most text similarity. Text preprocessing of the scraped and formalized job titles removed variabilities that would make association more challenging.

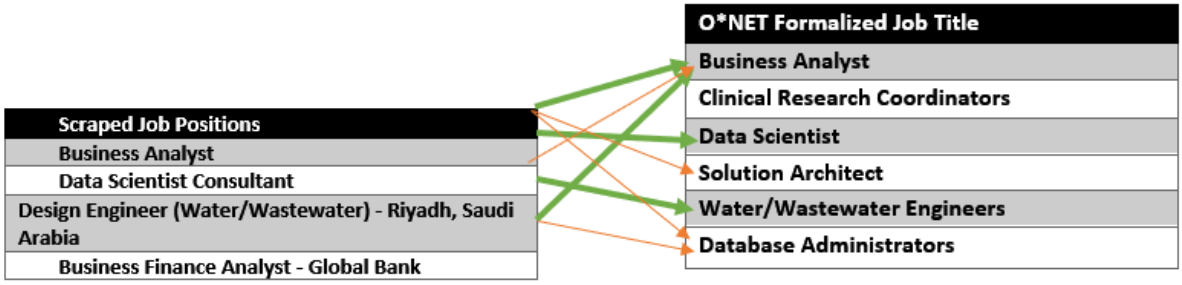


*Figure 5 Mapping from scraped unclean job titles to well-defined O*NET job titles*

Due to the extensiveness of the O*NET data, the majority of the scraped job titles was properly formalized, based on qualitative assessment of the mapped titles.

D. Job Description Sentence Classification: Skill or Task?

In a previous course assignment, we were asked to perform a topic modelling analysis on the current project dataset. Upon getting the results of the model and basing on the work presented by Kobyashi et al., it was noticed that the obtained topic models were obscured by terms that pertain to job position tasks. Tasks can be detrimental for a proper job market analysis as they are almost

always company-specific and would cloud important terms which would be found in sentences classified as skills, such as software names and/or languages.

For this reason, it was necessary to develop a classification model that would take care of classifying sentences/phrases in the dataset as “skill” or “task”. The model will be trained using data of approximately 65000 rows collected from Kobayashi et al. paper resources available on the paper site (Kobyashi et al., 2017). The data set was split into training and valid sample (60:40). Naturally, the training data was preprocessed and cleaned, but numerical were kept as they can represent required years of experience. After that, feature extraction was performed on the set in order to extract the features to train the model. 28 features were extracted in total.

The chosen features were specified to be:

- The Part of Speech (POS) of the first word in each element; ***category***
- Is the first word in an element unique to the “tasks” label?; ***boolean***
- Is the first word in an element unique to the “skills” label?; ***boolean***
- Is the last word in an element unique to the “tasks” label?; ***boolean***
- Is the last word in an element unique to the “skills” label?; ***boolean***
- TFIDF matrix (common features between training and test sample); ***numerical***

Using a Random Forest Tree Classifier, according to recommendations set by Kobyashi et al., the model was trained and then validated on the validation set. The out-of-sample accuracy was found to be approximately 94%.

The model was then used to predict whether each sentence in each job description is a skill or a task. Shown is an excerpt of the description sentences with the associated predicted labels:

Table 5 Classification Model Output

Job Description Sentence	Label
formulate , update manage customer services processes (customer services center , hospital delegates .)	**Task**
establish implement short long-range goals , objectives , policies , operating procedures .	**Task**
offer support services users coaching users within business maintain effective business relationships clients service providers (network members , insurance companies etc) to ensure services required organization delivered effective manner .	**Task**
develop maintain daily/weekly/monthly statistical reports to facilitate performance management customer services operations .	**Task**
profile excellent interpersonal communication skills quality focus customer oriented problem solving ability to handle stress analytical thinking & statistical skills negotiation skills computer literacy (ms word , ms excel , ms powerpoint) knowledge local market related health system knowledge related to medical terminologies would plus .	**Skill**
understanding insurance principles , business models would plus structured , analytical results oriented approach good command english .	**Task**
qualification/experience : minimum qualifications : bachelors degree marketing related field .	**Skill**
minimum experience : 6 years experience managing customer services functions .	**Skill**
minimum 2 years experience supervisory level terms people management .	**Skill**
experience similar environment tpa / medical insurance sector well would plus .	**Skill**

IV. Results

A. Topic Modelling

After splitting the sentences into skills and omitting sentences labelled as tasks, the filtered sentences were joined with their respective job position. Then for every job cluster, topic modelling was performed on its respective skill sentences. The topic models for three job titles are shown in the following tables, the rest of the job positions' topic models will be added to the appendix section.

Table 6 Latent Topics for Big Data Engineer Cluster

The Latent Topics of the Big Data Engineer cluster is:

Topic #1:
0.055*"data" + 0.014*"kafka" + 0.012*"expert" + 0.011*"hadoop" + 0.011*"work" + 0.011*"manag"
Topic #2:
0.017*"team" + 0.016*"process" + 0.012*"job" + 0.012*"data" + 0.012*"qualiti" + 0.011*"work"
Topic #3:
0.038*"data" + 0.022*"engin" + 0.017*"team" + 0.016*"oper" + 0.015*"big" + 0.015*"product"

Table 7 Latent Topics for Environmental Scientists and Specialists, Including Health Cluster

The Latent Topics of the Environmental Scientists and Specialists, Including Health cluster is:

Topic #1:
0.033*"environment" + 0.023*"work" + 0.022*"design" + 0.018*"support" + 0.016*"project" + 0.015*"develop"
Topic #2:
0.040*"environment" + 0.015*"requir" + 0.014*"health" + 0.014*"compani" + 0.013*"manag" + 0.013*"project"
Topic #3:
0.030*"environment" + 0.027*"project" + 0.019*"work" + 0.016*"design" + 0.016*"requir" + 0.015*"health"

Table 8 Latent Topics for Interior Designers Cluster

The Latent Topics of the Interior Designers cluster is::

Topic #1:
0.040*"design" + 0.026*"engin" + 0.019*"respons" + 0.014*"util" + 0.014*"wet" + 0.014*"key"
Topic #2:
0.044*"design" + 0.018*"riyadh" + 0.016*"engin" + 0.016*"year" + 0.016*"lead" + 0.014*"draw
Topic #3:
0.053*"design" + 0.025*"engin" + 0.022*"requir" + 0.017*"work" + 0.014*"electr" + 0.014*"draw"

An interactive dashboard was created to visualize the topic clusters for a given job title. Here, the Data Scientist position is represented with 5 latent topics. By analyzing the relevant terms on the right, we can also qualitatively identify the following topic clusters:

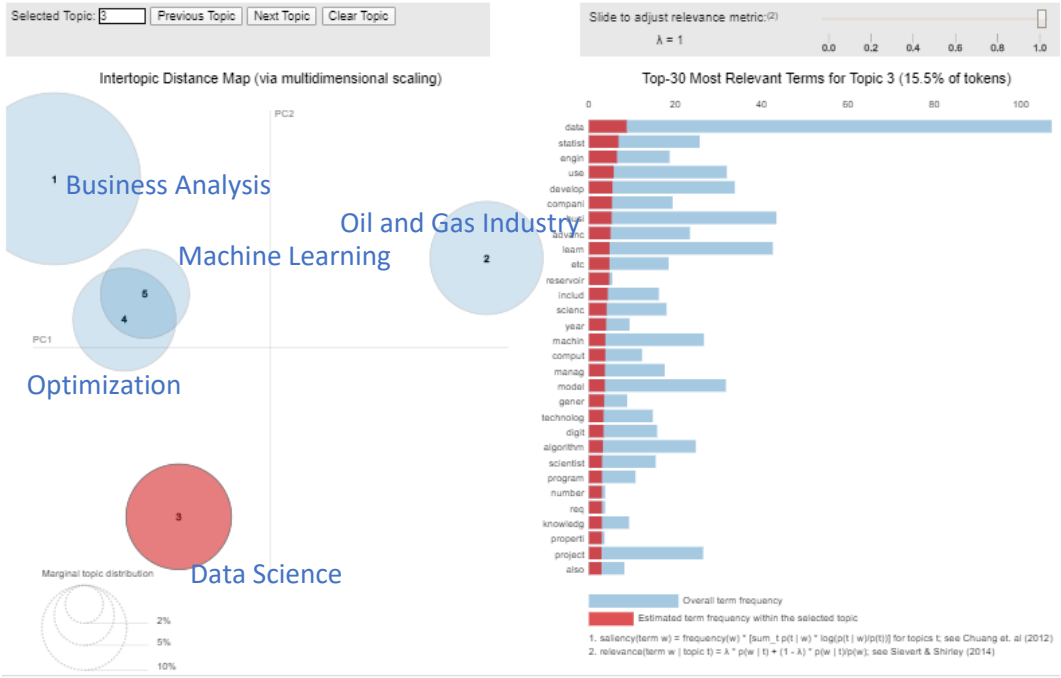


Figure 6 Generated Dashboard to visualize LDA Clustering and most relevent terms per cluster

B. Job Frequency

1) Per Career Cluster (High Level Clustering)

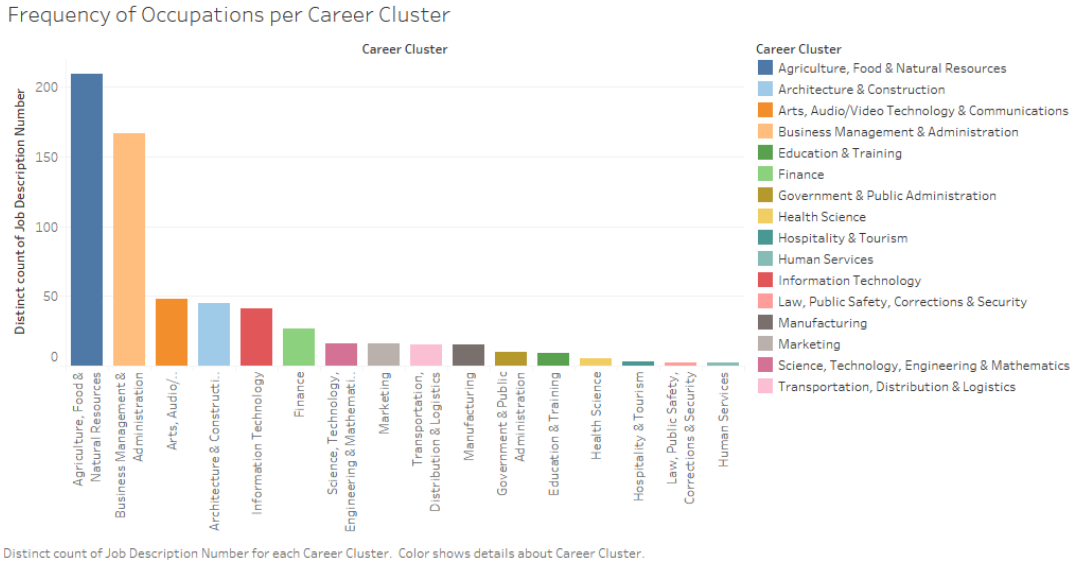


Figure 7 Frequency of Occupations per Career Cluster

2) Per Career Pathway (Medium Level Clustering)

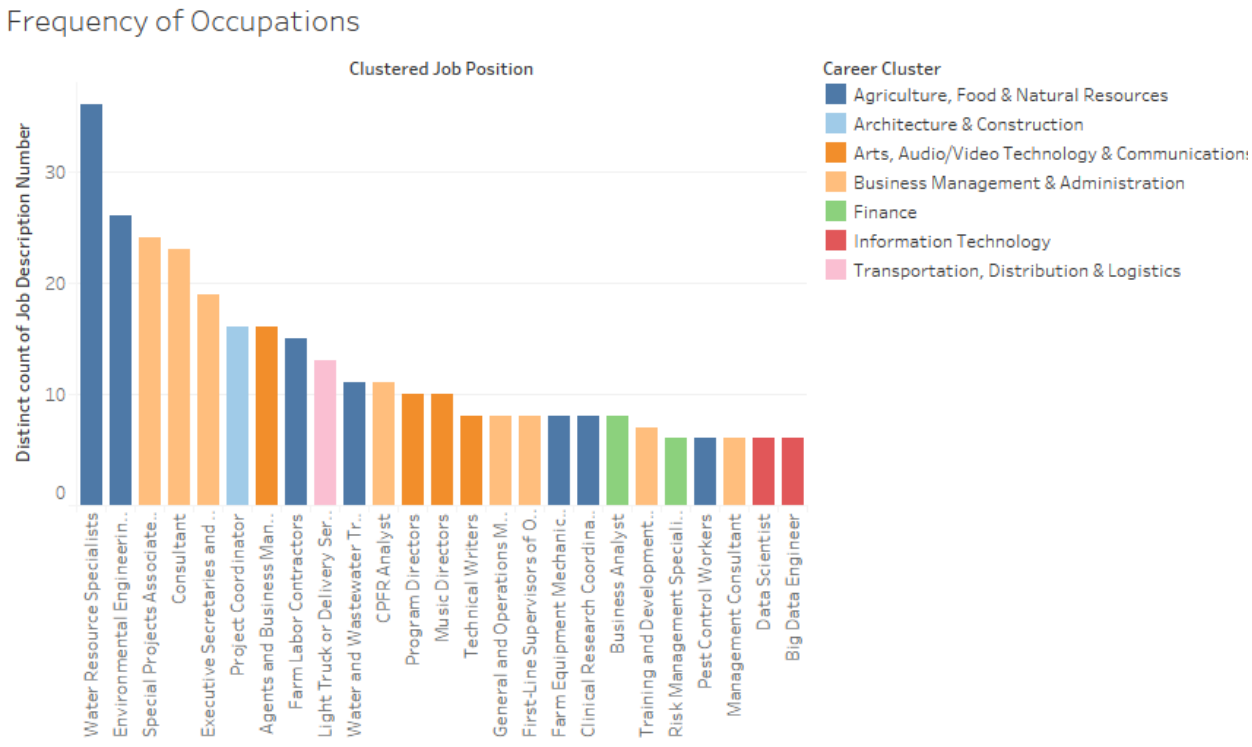


Figure 8 Frequency of Occupations per Career Pathways

3) Per Job Title (Low Level Clustering)

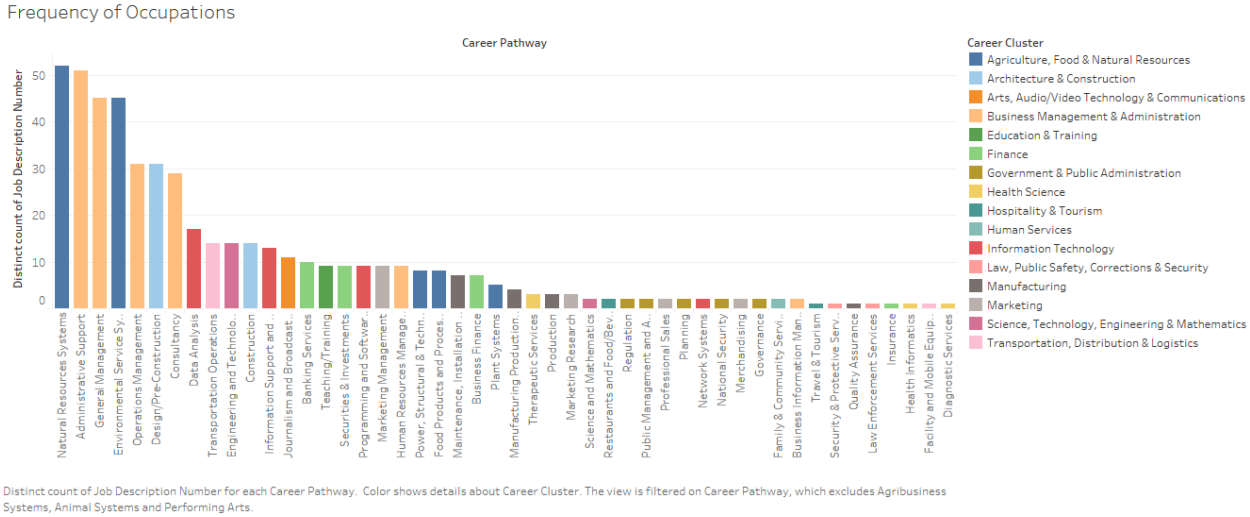


Figure 9 Frequency of Occupations per Job Title

The most in-demand job positions, it seems, are in the environmental, petrochemical, and administrative jobs.

C. Years of Experience

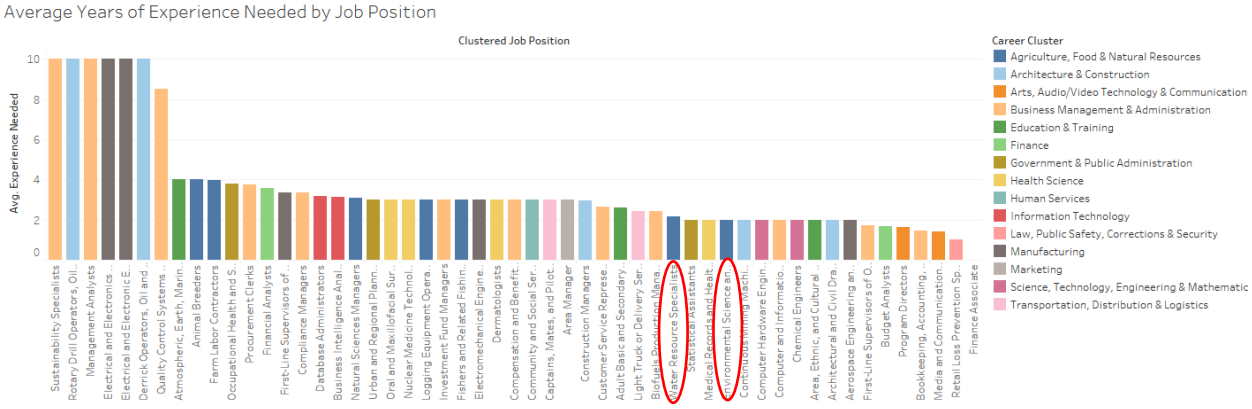


Figure 10 Average Years of Experience Needed by Job Position

Jobs in Management, Manufacturing, and Architecture recorded the highest average of required years of experience with up to 10 years required on average for some positions in these fields. The jobs that required the most years of experience were Management Analysts, Sustainability Specialists, Electric Technicians, and Oil Drill Operators. Most of these job positions require extensive handy work and would therefore require that the individual have had extensive experience.

D. Job Market Trend: Time Series

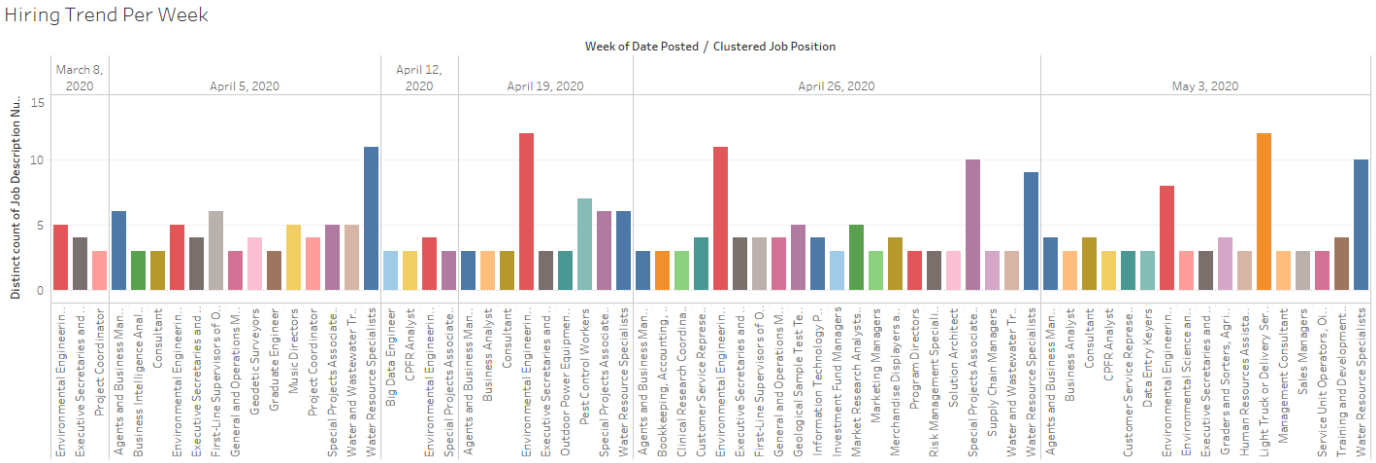


Figure 11 Time Series of Job Market Demand per week

The demand for Environmental engineers has been increasing in the last few weeks. The demand for water resource specialists is also experiencing peaks in a weekly basis in comparison to other jobs.

V. Discussion

In our analysis, we have shown that the Saudi job market is in need for highly skilled individuals and technicians, namely in the power and services sector. Petroleum Engineers, Water Resources Specialists, Environmental Engineers, and manufacturing technicians are all in high demand. At the same time, Environmental and Water engineering openings are, on average, open for individuals with up to 3 years of experience only. Administrative assistants with a few years of experience are also in high demand. This signifies that fresh graduates/seniors have a higher chance of getting incorporated into the Saudi job market if they have such profiles. On the contrary, hiring managers in the Oil and Gas industry are on the lookout for Petroleum Engineers and Drill Operators with up to 10 years of experience.

Moreover, using our time-series variation of job demand, policy makers will have access to a "live-feed" of the offered job position, and will be able to draw trends describing rates of change in the demand for each occupation. Policy makers may also utilize the time-series to see the effect of certain measures they take on the capacity of a certain industry to hire new people. For example, they may notice that a newly introduced tax on oil companies resulted in a sharp decrease in the available job opportunities in the sector, and they would therefore revoke the tax.

VI. Conclusion

As a conclusion, we have demonstrated through our analysis, that, with the proper text processing and modelling techniques, one can really capitalize upon open-source data available on platforms like LinkedIn so as to generate insights on the state of the job market at a given moment in time. After aggregating job titles into uniform tags and using feature engineering and Random Forest Classifier were able to extract only key information pertaining to skills and not job tasks from each job description. Then, using LDA, we were able to really go deeper into every job position and to see acquire sub-clusters for each job position by visualizing them in a dynamic dashboard. Then using NER, we extract information regarding the sought years of experience for each job cluster. Also, a time-series was developed using the extracted "date posted" feature on LinkedIn. We agglomerated all the above-mentioned insights into clear recommendations in the form of job titles that are currently in demand and those that do not require a lot of work experience, for policy makers to act accordingly.

One of the most challenging aspects of the project was the data collection process. The selenium app that was developed relies heavily on the html tags configuration on LinkedIn. LinkedIn sometimes enforces some minor changes in the html configuration of its pages which would render our scraper helpless and dysfunctional. To fix this issue, the scraper source code is manually edited so as to match the html configuration on LinkedIn.

Finally, this project will represent the steppingstone for a full-pledged capstone project to be carried on during the Summer semester 2020. The next steps of our study will aim to expand the survey of the job market to more than one Arab country. Moreover, the capstone will aim to dive deeper into the time-aspect of job market trends, and to provide more insight from job descriptions that can also be collected from more resources if needed.

VII. Acknowledgements

This project really reinforced the course learning experience and expanded technical capabilities!

I would like to thank Dr. Wael Khreich for his continuous one-on-one support and for his valuable feedback. This project, and others to come, would not have been possible without the efforts of Dr. Khreich.

VIII. References

Best Practices for Your Mobility Policies. Retrieved from https://mobilityexchange.mercer.com/Insights/article/Best-Practices-for-Your-Mobility-Policies

Karakatsanis, I., Alkhader, W., Maccrory, F., Alibasic, A., Omar, M. A., Aung, Z., & Woon, W. L. (2017). Data mining approach to monitoring the requirements of the job market: A case study. Information Systems, 65, 1–6. doi: 10.1016/j.is.2016.10.009

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Hartog, D. N. D. (2017). Text Mining in Organizational Research. *Organizational Research Methods*, *21*(3), 733–765. doi: 10.1177/1094428117722619

O*NET® 24.2 Database. Retrieved from https://www.onetcenter.org/database.html

Robertson, A. (2019, September 10). Scraping public data from a website probably isn't hacking, says court. Retrieved from https://www.theverge.com/2019/9/10/20859399/linkedin-hiq-data-scraping-cfaa-lawsuit-ninth-circuit-ruling

Smith, D., & Ali, A. (2014). Analyzing computer programming job trend using web data mining. Issues in Informing Science and Information Technology, 11, 203-214. Retrieved from http://iisit.org/Vol11/IISITv11p203- 214Smith0494.pdf

The Future of Jobs and Skills in the Middle East and North Africa: Preparing the Region for the Fourth Industrial Revolution. (n.d.). Retrieved from https://www.weforum.org/reports/the-future-of-jobs-and-skills-in-the-middle-east-and-north-africa-preparing-the-region-for-the-fourth-industrial-revolution