CRACOW UNIVERSITY OF ECONOMICS

Dwijendra Nath Dwivedi

M. Phil. In Development Research

Bachelor in Computer Science and Engineering

PhD Dissertation:

Analysis of the employer's expectations towards competencies of candidates for employment in the Banking Sector in The UAE using Text Mining Approach

SCIENTIFIC SUPERVISOR:

PROF. DR HAB. PAWEL LULA

KRAKÓW 2023

Scientific Discipline: Management and Quality Sciences

Dyscyplina naukowa: nauki o zarządzaniu i jakości

Acknowledgements

I am deeply grateful to my mother, who has been a constant source of inspiration, love, and affection that I miss every day. My father always encouraged us to make a difference and make him feel proud. My siblings have also been a big support in my life. I also miss my guruji, who always encouraged me to strive for excellence in both the physical and spiritual realms.

I am also thankful to my wife, Pragya, who supported me throughout my PhD journey by taking care of the children and their studies, allowing me to focus on my studies in Kraków and my thesis and articles.

I am immensely grateful to my scientific director, Prof. Dr hab. Pawel Lula, for the exceptional support he has provided throughout my PhD program. His guidance has been invaluable in shaping my doctoral thesis.

I also want to express my gratitude to the Inter-Faculty Doctoral Committee for their advice on refining my doctoral proposal. I am grateful to all the teachers I had the opportunity to learn from during my program. Finally, I want to thank all my friends for their motivation and support.

Abstract

A major factor in the company's competitive advantage is the alignment of employee competencies with the organization's focus. One challenge in this regard is that the adoption of new technologies and practices requires that employees in the organization develop new skills at a faster pace than before. The constant evolution of technology adoption within organizations has an impact on the labor market. Competency models are very useful tools for ensuring that HR systems facilitate and support a business's strategic goals. This increases the potential for placing the right people in the right jobs. At the decision level, institutions review the succession provisions to see how many individuals who have been employed fall into the high-potential category (McIlvaine, 1998). The skills available are becoming obsolete. The adoption of new practices leads to the creation of a new skill. Organizations need to assess the skills needed to do the work, identify the gap, and create a plan to bridge the gap. Thus, it is important to develop opportunities to identify employer skill demand and skill supply (Lula et al, 2019).

The main objective for the thesis is to perform the analysis of competency gap in the UAE banking sector; here we propose to use the competency schema proposed by Lula et al (2019) where authors defined as a set competency and a set of relations between them, together with the information about the importance of every competency and the importance of every relationship between any two competencies. We propose to extend the same concept to identify the competency schemas for banking sector in the UAE. To solve these questions, we will need to define the competency schemes in the local market and identify the relationship between schemas. To perform existing competency analysis, I. e., identification of existing taxonomy and presentation of research works related to analysis of competencies, especially in banking sector in the United Arab Emirates. Conduction of analysis and presentation of research findings allows to

- *present main features of the UAE economy, especially the banking sector*
- *identify crucial competencies expected in the banking sector in the UAE*

- *develop a methodology for analysis of expected competencies*
- *building a network model based on bipartite graphs allowing to describe relationships between various aspects of labor market*
- *build a software tool for performing a complete analysis of supply side and demand side of this labor market*

Chapter 1 presents and discusses Understanding of the UAE banking sector and the labour market. A review of the banking sector shares details about the public and the private bank structure in UAE. Also, some labour market statistics are shared.

Chapter 2 focuses on the competencies and their analysis. Organizations need to assess the competencies that are needed to do the job, identify the gap, and create a plan to bridge the gap. It thus becomes significant to develop the possibilities of identifying the employer demand for competencies and the supply of competencies.

Chapter 3 provides the theoretical background and illustrates methodological aspects of text mining to be used to analyse the Job discrepancy database.

Chapter 4 performs analysis of competency demand for banking sector in the UAE using text mining methods. Text mining is an artificial intelligence (AI) technology that uses natural language processing (NLP) to extract meaningful information from large amounts of textual information. One of the methods for the same could be Latent Dirichlet Allocation (LDA). LDA is a statistical and graphical model which is used to obtain relationships between multiple documents in a corpus.

Chapter 5 provides the theoretical background and introduce graph-based models in text mining of job dependencies. Graphs here represent co-occurrence of words in text segments or in documents. They are useful for identification of keywords and key-phrases.

Chapter 6 highlights the empirical result of competency analysis for banking sector in the UAE. Also, we text the network statistics for demographics of job postings.

Chapter 7 focuses on the conclusions, recommendations, and further research. Conclusion encompasses the discussion on empirical findings, limitations of this study, recommendations for future research and conclusion.

Contents

Introduction

Because of the changing environment, organizations need to keep track of the effectiveness and competent labor market. New methods are necessary for the same study and the literature shows the significance of competences on labor markets. Literature review fails to agree on the competency framework for banks or other industries. A scheme of scientific competence is needed for the banking sector in the United Arab Emirates. There is no ontology/text mining-based study that examines the representation of the domain knowledge concerning competences crucial for employers, employees and candidates looking for jobs in banking sector in the UAE. Furthermore, there is a gap in the prioritization of skills in the banking sector in the UAE. There is a no competency gap assessment tool that can perform a gap assessment in the fast-changing banking market in UAE.

Based on the above literature study, we discovered #3 main research gaps justifying our current study. The dissertation idea is made up of two layers of literature review. It starts with the study of competencies in general. It is followed by its study of such in the banking sectors. The main objective of this dissertation is to develop a software to perform the competency gap assessment in the UAE market mainly focusing on the banking sector.

Main Hypothesis:

- Hypothesis 1: Exploratory text analysis of job descriptions can be used for identification of crucial competencies expected in the banking sector in the UAE.
- However, ontology-based approach allows to get better results than corpus-based one.
- Hypothesis 2: Competency schemas allow for an in-depth description and examination of relationships between expected competencies.
- Hypothesis 3: Network models are useful for description of various aspects of labour market in banking sector in the UAE.

Within the main objective, we have intermediate objectives such as

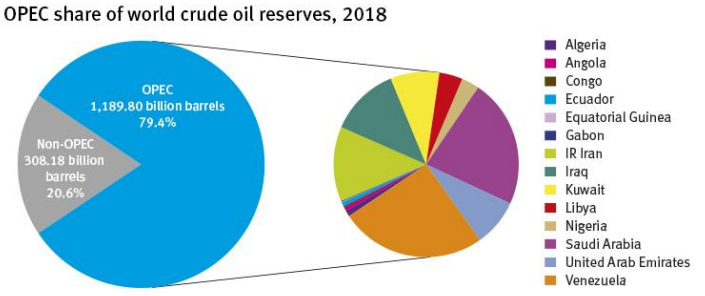
- Intermediate Goal1: Understanding of the UAE banking sector and the labor market
    - Understanding of the UAE as a country with the role of public and private banking sectors in economy.
- Intermediate Goal2: Identify competency schemas for banking industry
    - Presentation of research works related to analysis of competencies from different industries, especially from the banking sector.
- Intermediate Goal3*:* Application of corpus-based approach in analysis of published job description in banking sector
    - To detect the key competencies and relationship between them. The analysis will be performed with the use of the Latent Dirichlet Allocation method that is a probabilistic model where each document is assumed to have a combination of topics. It assumes that each document comprises a small number of topics and that each word can be attributed to one of the topics contained in the document.
- Intermediate Goal4*:* Application of ontology-based approach in the analysis of competencies expected in the banking sector in the UAE:
    - presents a competency schema for the banking sector in UAE
    - Ontology-based approach tries to describe a given domain using labelled graph-based models. Objects are represented by the graph's nodes and relationships between objects are described by the graph's edges. We use the ontology-based approach in the analysis of published job descriptions from the banking sector to detect the key competencies and their relationships.
- Intermediate goal 5: Importance of key competencies.
    - Highlight the key competencies for banking sector in the UAE.

1. Chapter 1: United Arab Emirates Labor Market Analysis

1.1 Introduction:

The UAE is a Middle Eastern country located in the western portion of the continent of Asia and a federation of seven states. Abu Dhabi city is the largest emirates and the capital of the UAE. The country has a widespread coastline along the Persian Gulf and the Sea of Oman and 4 hours ahead of GMT. It is surrounded by Saudi Arabia to the west and south and by Oman to the east and northeast. The country's economy is driven by its two largest cities, Dubai, and Abu Dhabi.

The UAE is home to over 200 nationalities where its citizens, the Emiratis constitute over 20% of the total population, making the UAE home to one of the world's highest percentage of immigrants. The UAE also ranked 1st globally in peaceful coexistence among nationalities, with residents from over 200 countries, according to the report of the UN International Organization for Peace, Care and Relief for 2014. As per UNDP 2014 report on Human development indicators, the UAE finds a place in the segment of countries with very high HDI (0.835). According to the same report, the average life expectancy of citizens in the UAE increased to 76.8 years as compared to 76.7 years in the 2013 report, while the average number of years of schooling for citizens has now reached 13.3 years, compared to 12 years in the last report. The UAE is a major oil and gas exporting nation, it holds almost 97.8Bn barrels of proven oil, it is ranked as the sixth largest proven reserves crude oil within OPEC nations.



OPEC proven crude oil reserves, at end 2018 (billion barrels, OPEC share)

Venezuela	302.81	25.5%	Kuwait	101.50	8.5%	Algeria	12.20	1.0%	Gabon	2.00	0.2%
Saudi Arabia	267.03	22.4%	UAE	97.80	8.2%	Ecuador	8.27	0.7%	Equatorial Guinea	1.10	0.1%
IR Iran	155.60	13.1%	Libya	48.36	4.1%	Angola	8.16	0.7%			
Iraq	145.02	12.2%	Nigeria	36.97	3.1%	Congo	2.98	0.3%			

Source: OPEC Annual Statistical Bulletin 2019.

Graph1: Oil reserves at OPEC member countries (2018) Source - OPEC

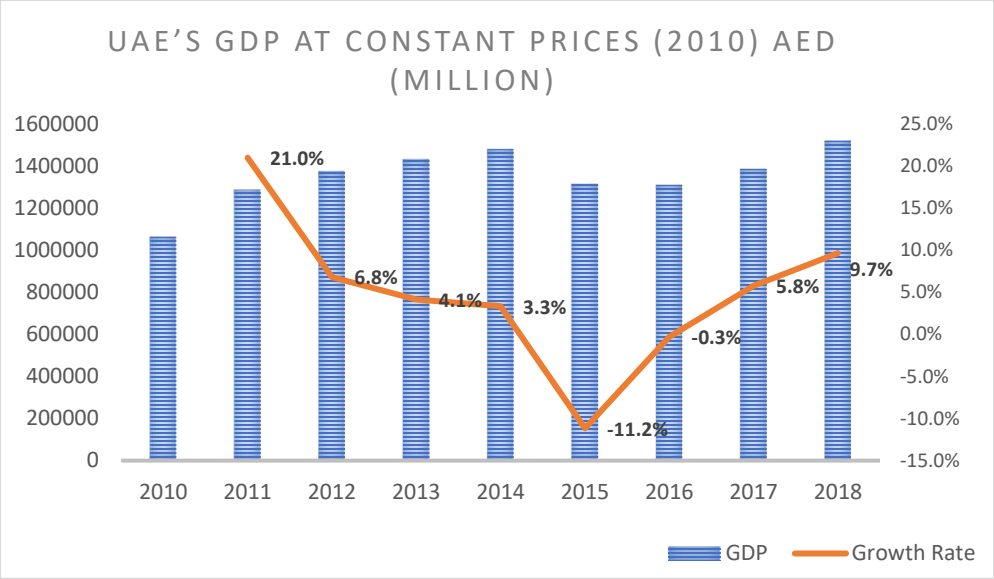
1.2 Economic Indicators:

The UAE 2021 Vision recently taken by the Government looks at innovation and knowledge as the key drivers of the economy. In the 1950s the UAE's economy was dependent on fishing and a declining pearl industry when oil discovery was not made. The country's society and economy have been transformed after oil and gas discovery was made. The pro Oil era, the region's economy was driven by mainly nomadic agricultural, date palm farming, fishing, pearling and seafaring. Post oil discovery, the economy has been expanding mainly by the following sectors:

- extraction of crude oil and natural gas
- repair services
- wholesale and retail trade
- real estate and business services
- construction and manufacturing
- Banking and financial services
- Gold and diamond trading

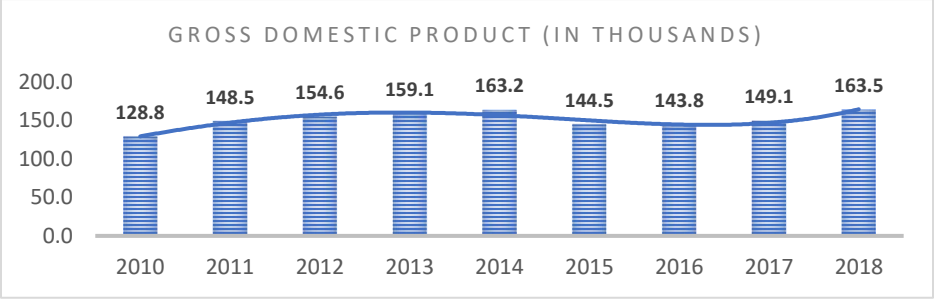
1.2.1 The Gross Domestic Product (GDP) is one of the important economic variables that helps recognize the economic presentation of the nation. It represents the total dollar value of all goods and services produced over a specific time, often referred to as the size of the economy. As per the *Federal competitiveness and statistics authority UAE,* in 2018, the UAE's GDP at current prices was AED 1.52 trillion. Its GDP at constant prices (2010) shows a 35.5% increase between 2010 and 2018.

As per *World Bank statistics (2018), The UAE* is ranked 29th in the global GDP raking with 414 billion. As per *International Monetary Fund (IMF)* consultation report with the United Arab Emirates, growth forecast is projected at around 3.7 percent for 2019–20. Non-oil growth is predicted to rise to 3.9 percent in 2019 and 4.2 percent in 2020.



Graph2: Federal Competitiveness and Statistics Authority UAE's GDP at constant prices (2010) between 2010 and 2018.

GDP Per Capita: As per Federal competitiveness and statistics authority the UAE, GDP per capita has been growing for most, if the years, except 2015 when it shrank by 11% due to oil process shock. It grows sharply by 10% in 2018. Average GDP per capita growth rate for last 10 years has been 3%.



Graph3: Federal Competitiveness and Statistics Authority UAE's GDP Per capita

1.2.2 Sector Contribution to GDP: In 2018 economic activity that contributed most to the overall GDP (current prices) was mined and quarrying (including crude oil and natural gas) that contributed almost 29.5% of GDP. Wholesale and retail trade, repair of motor vehicles and

motorcycles was the second largest contributor to GDP. Financial services have contributed to 8.6% and is the third largest sector.

Economic sector	Sector's contribution to the GDP for 2017 (in per cent)
Extractive Industries (include Crude Oil & Natural Gas)	29.50%
Wholesale and Retail Trade & Repair of Motor	11.70%
Financial Services and Insurance Activities	8.60%
Building Construction	8.40%
Transformative Industries	8.30%
Public Administration and Defense & Compulsory Social Security	5.80%
Real Estate Projects	5.70%
Transport and Storage	5.40%
Gas and Water, Electricity	3.20%
Information & Communications	2.90%
Professional, Scientific &Technical Activities	2.60%
Other sectors	8.00%

Table1: contribution of the economic sectors in the GDP for 2017 at real prices of 2010

https://www.government.ae/en/about-the-uae/economy

1.2.3 Federal Budget:

The government of the United Arab Emirates had a 2-year, zero-deficit budget. The amount of the budget has been raised more than 300 times since the first budget was allocated in 1971.

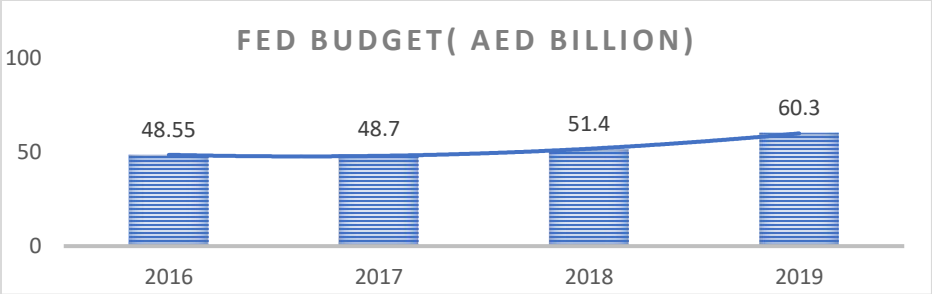


Table 2: Federal budget from 2016-2019.

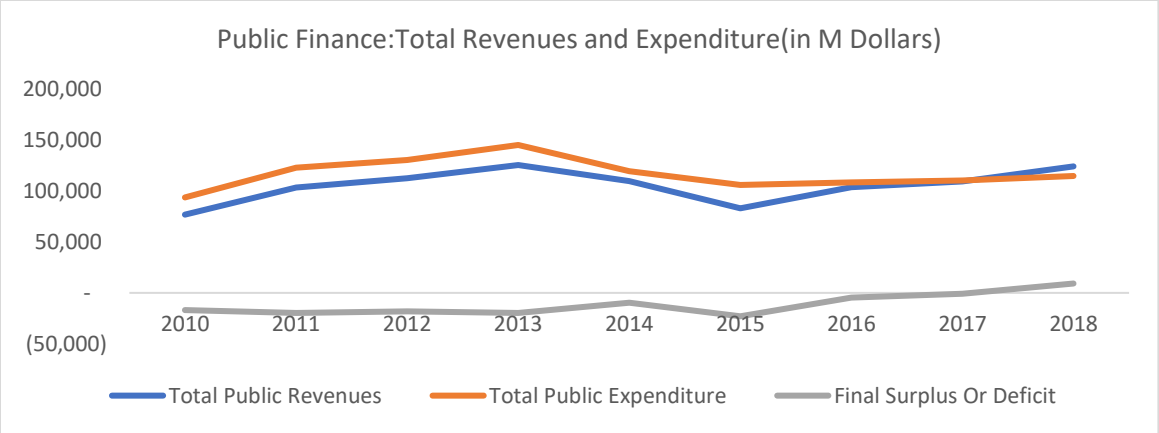
https://www.government.ae/en/information-and-services/finance-and-investment/federal-finance

1.2.4 Public Finance: Total Revenues and Expenditure:

The UAE economy became surplus in 2018 by 9.18 B (Dirhams) after having almost 10 years of deficit. Revenue has grown by 13.3% in 2018 whereas the expenses have grown by 4.2% resulting in a net surplus. Another highlight is the oil price shock that caused a major revenue dropped by almost 33.9% in 2015 from the high of 2013.

Public Finance:Total Revenues and Expenditure, 2010 - 2018 (Million Dollars) United Arab Emirates									
Item	2010	2011	2012	2013	2014	2015	2016	2017	2018
Crude oil Reveneus	46,230	71,460	76,045	79,849	69,083	37,558	23,623	39,451	44,715
Other Revenues	30,563	31,975	36,337	45,647	40,739	45,433	80,203	69,982	79,319
Total Public Revenues	76,793	103,435	112,382	125,496	109,822	82,991	103,826	109,433	124,034
Total Public Expenditure	93,679	123,042	130,524	145,168	119,451	105,988	108,436	110,184	114,846
Final Surplus Or Deficit	(16,886)	(19,607)	(18,142)	(19,672)	(9,629)	(22,997)	(4,610)	(750)	9,188

Table 3: Public Finance: Total Revenues and Expenditure, 2010–2018 Source: Federal Competitiveness and Statistics Authority UAE



Graph4: Financial surplus or deficit trend of public finance (Source: Federal Competitiveness and Statistics Authority UAE)

The UAE has been trying to grow non-oil revenue for last few years, and we can see a higher share of non-oil sectors for the same.

Public Revenues, 2010 - 2018 (Million Dollars)									
Item	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Tax Revenues	**9.3%**	**5.2%**	**8.1%**	**5.6%**	**6.8%**	**7.1%**	**5.9%**	**5.5%**	**5.5%**
Customs	2.9%	2.5%	2.6%	2.1%	2.6%	3.7%	2.7%	2.3%	2.3%
Other	6.4%	2.7%	5.5%	3.5%	4.2%	3.4%	3.2%	3.2%	3.2%
Non Tax Revenues	**90.7%**	**94.8%**	**91.9%**	**94.4%**	**93.2%**	**92.9%**	**94.1%**	**94.5%**	**94.5%**
Oil and Gas	60.2%	69.1%	67.7%	63.6%	62.9%	45.3%	22.8%	36.1%	36.1%
Enterprise Profits	7.6%	6.0%	5.1%	5.6%	4.9%	6.0%	44.9%	32.9%	32.9%
Others	22.8%	19.8%	19.1%	25.2%	25.4%	41.7%	26.4%	25.6%	25.6%
Total Revenues	76,793	103,435	112,382	125,496	109,822	82,991	103,826	109,433	124,034

Table 4: Public Revenues 2010–2018 (Source: Federal Competitiveness and Statistics Authority UAE)

1.2.5 Annual trade surplus:

The UAE has a structurally positive trade balance economy. Main imports have been pearls, gold, diamond and other valuable metals and stones; equipment, sound recorders, reproducers and parts and transport automobiles. United Arab Emirates' main trading partner is India (14 percent of total exports and 17 percent of imports). Others include Japan, South Korea, China, United States and Iran. Oil products are by far the largest item of exports, followed by gold, diamond and jewelry.

United Arab Emirates Trade Surplus, 2010 - 2018 (Million Dirhams)									
Economic Variables	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Exports of Goods & Services	875,260	1,160,025	1,379,075	1,440,518	1,474,019	1,326,700	1,324,400	1,410,400	1,427,700
Imports of Goods & Services	861,976	1,046,489	1,163,105	1,225,449	1,324,655	1,268,100	1,286,800	1,268,900	1,226,000
Surplus	13,284	113,536	215,970	215,069	149,364	58,600	37,600	141,500	201,700

Table 5: United Arab Emirates Trade Surplus, 2010 - 2018 (Million Dirhams) (Source: Federal Competitiveness and Statistics Authority UAE)

1.2.6 Emirates Securities Market Index:

DFM, ADX and NASDAQ Dubai are the three stock exchanges in the UAE. Abu Dhabi Securities Exchange (ADX) lists mostly the UAE companies. NASDAQ Dubai was set up to trade international stocks. Dubai based Dubai Financial Market is a stock exchange established in 2000 where ADX is based out of Abu Dhabi. DFM and ADX are both governed and regulated by the Securities and Commodities Authority (SCA). The UAE's financial markets deal primarily in equities, securities, bonds, futures, mutual funds, commodities, currencies, metals, stones, derivatives, and Sukuk (Islamic bonds).

The Stock Market Capitalization to GDP ratio is calculated by Sum of (The Stock Market Capitalization of the Abu Dhabi Securities Exchange (ADX) and of the Dubai Financial Market (DFM) as of End of the year) divided by GDP of the UAE. The stock market capitalization-to-GDP ratio is a quantitative ratio used to estimate whether an overall market is undervalued or overvalued compared to a historical average. If the ratio falls between 50 and 75%, the market can be said to be modestly undervalued. It is modestly overvalued if it falls within the range of 90 and 115%.

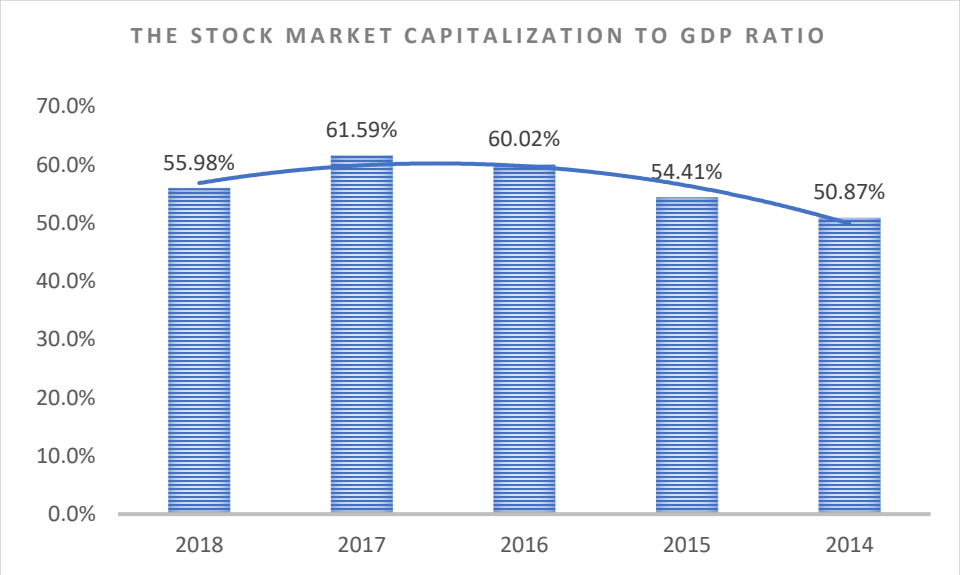
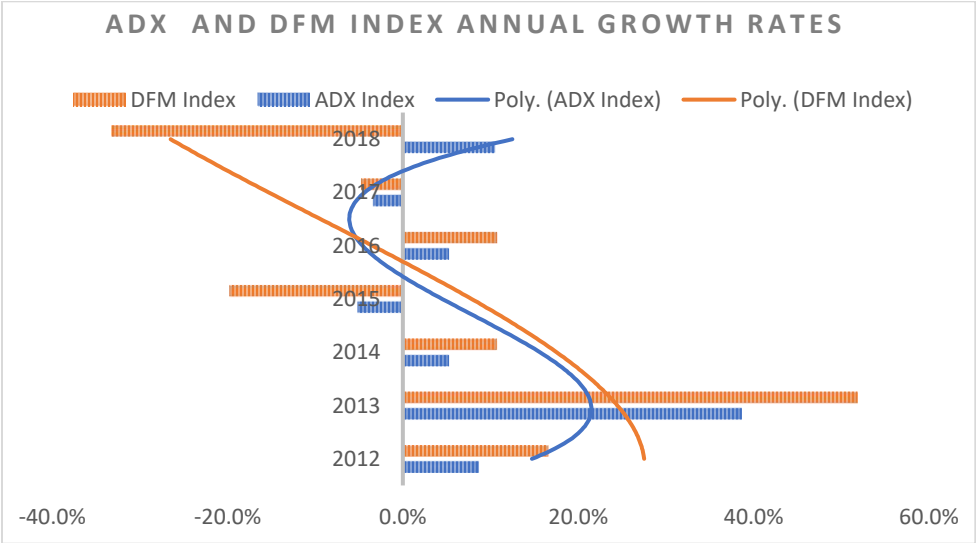


Table 6: Stock Market Capitalization to GDP Ratio (Source: Securities & Commodities Authority)

ADX has increased by 10% in 2018 but the DFM index has declined by 33% in the same year. . The number of companies which have been listed for trading has been in the market.



Graph 5: ADX and DFM index Annual Growth Rates from 2012 to 2018 Source: Securities & Commodities Authority

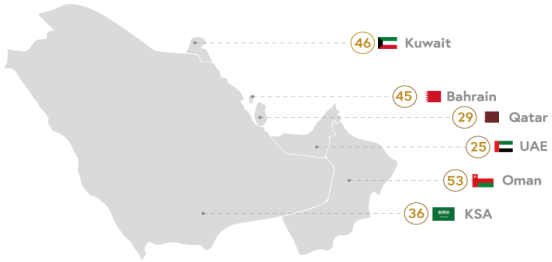
.

Emirates Securities Market Index For 2011 - 2018									
Item	**Unit**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Total Number of listed Companies*	**Number**	135	129	127	125	128	129	134	137
ADX Index	**Point**	2,402.28	2,630.86	4,290.30	4,528.93	4,307.26	4546.37	4,398.44	4,915.07
DFM Index	**Point**	1,353.39	1,622.53	3,369.81	3,774.00	3,151.00	3,530.88	3,370.07	2,529.75
Total Market Capitalization**	**Billion AED**	447.13	478.52	714.32	753.08	715.57	787.00	854.34	851.44
Total Traded Volume**	**Billion Shares**	40.98	56.82	178.69	216.80	117.92	131.11	107.63	58.69
Total Traded Value**	**Billion AED**	56.72	70.65	244.51	525.23	202.29	175.30	159.60	95.48
Total Number of Trades**	**Number**	728,103	880,039	1,894,090	3,264,736	1,954,216	1,643,811	1,414,974	933,625
Source: Securities & Commodities Authority									

Table 7: Federal Competitiveness and Statistics Authority UAE

1.2.7 Ease of Doing Business and Global Competitiveness Report and Ranking:

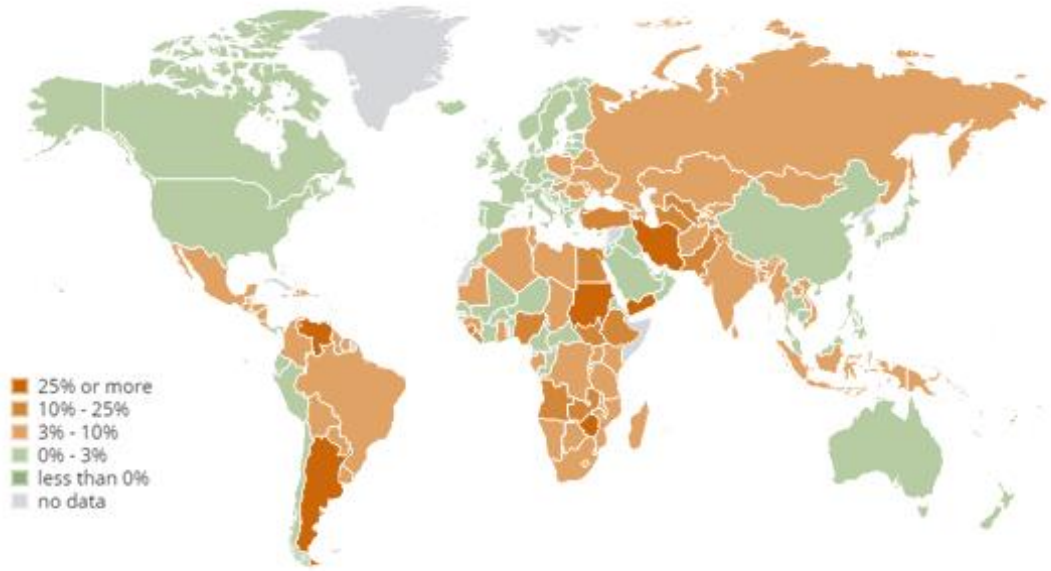
The ease of doing business ranking is an index shaped jointly by Simeon Djankov and Gerhard Pohl, two leading economists at the World Bank Group. A higher ranking of doing business means the governing operational environment is more reassuring to the starting and operation of a local firm. The UAE is a leader in the Middle East and Arab region in the World Bank's ease of doing business ranking 2020 with a 16th position in the global ranking of 190 countries. The country was ranked 11th position in 2019. Despite a dip in the overall ranking, the UAE continues to keep its score high in key areas of the global ranking. As per the Global Competitiveness Index 4.0 established by World Economic Forum that ranks 141 countries across the 12 pillars (Institutions, Infrastructure, ICT adoption, Macroeconomic stability, Health, Skills, Product market, Labor market, financial system, Market size, Business dynamism and Innovation capability), the UAE is ranked 25th. The 12 pillars include 103 indicators (70 % of weight is based on hard data and 30% is based on surveys.



Graph 6: World Economic Forum Global competitiveness ranking GCC countries

1.2.8 Inflation statistics:

In economics, Inflation is the decline of purchasing power of a given currency over time. The most common measure of inflation is the inflation rate. That is the annualized percentage change in a general price index made up of a basket of goods and services. In 2017, the inflation rate of the United Arab Emirates was at 2 percent, but it jumped to the sharp increase of over 3.1 percent in 2018. Overall UAE has witnessed low inflation statistics over the year.



Graph 7: IMF world economic inflation outlook 2019

https://www.imf.org/external/datamapper/PCPIPCH@WEO/OEMDC/ARE/SAU?year=2020

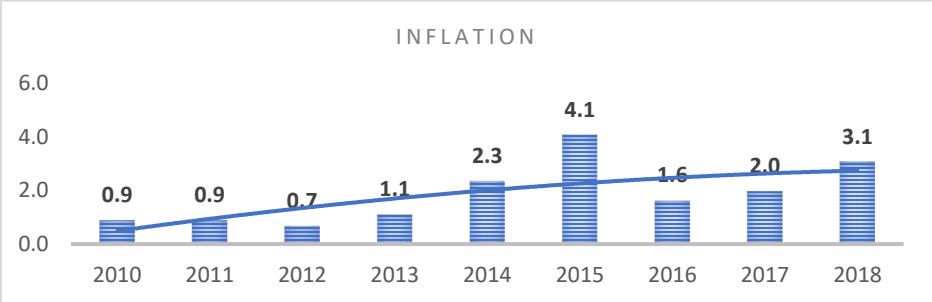


Table 8: inflation statistics for UAE since 2010 :2018 (Source: Federal Competitiveness and Statistics Authority UAE)

1.2.9 Tourism & Hotel Industry:

As per the government of the UAE (https://www.government.ae/en/information-and-services/visiting-and-exploring-the-uae/travel-and-tourism) In 2016, the *direct* role of the travel and tourism sector to the UAE's GDP was AED 68.5 billion (USD 18.7 billion) equivalent to 5.2 per cent of the total UAE GDP. The *total* contribution of the travel and tourism sector to the UAE's GDP was AED 159.1 billion (USD 43.3 billion) which is 12.1 per cent of GDP. The direct contribution of the travel and tourism sector supported 317,500 jobs in the UAE, which is 5.4 per cent of total employment.

2017 has seen a 7.7 % increase in the total number of guests, though the length of stay of guests has decreased a little bit. Hotel industry has seen a robust growth that increased the availability of rooms by 4.2%.

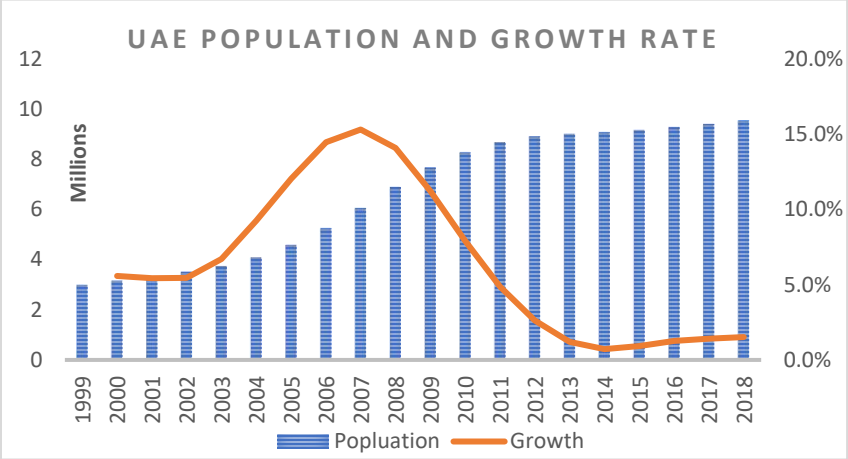
Hotel Establishments Main Indicators, 2017			
Indicator	**2016**	**2017**	**% Growth Rate**
Actual Guest Arrivals (No.)	22,868,489	24,633,790	7.7
Guest Nights (No.)	73,854,531	78,179,269	5.9
Length of Stay (Avg)	3.23	3.17	-1.8
Available Rooms (No.)	155,704	162,225	4.2
Occupancy room Rate (%)	75%	76%	1.5
Total Revenue	31,630,805,379	31,747,362,818	0.4
Room Revenue	19,199,591,199	19,286,793,192	0.5
Food & Beverages	11,515,955,742	11,474,166,341	-0.4
Other Revenue	915,258,438	986,403,284	7.8
ARR	455	436	-4.2

Table 9: Hotel Indicators 2016 & 2017, the UAE (Source: Federal Competitiveness and Statistics Authority UAE)

The UAE had been one of the top tourism destinations due to Economic and political stability, strategic location connecting the East and the West and low crime rate - the rate of premeditated murders in the UAE is 0.3 per cent for every 100,000 people as of January 2018.

1.3 Population and Gender statistics

As per *world bank statistics*, United Arab Emirates population as of 2018 stands at 9.6 Million. From 2010 to 2018, the UAE's population increased by a total of 12.6% and from 2017 to 2018, by a total of 1.5%. Population in the world is currently (2019-2020) growing at a rate of around 1.08% per year (down from 1.10% in 2018, 1.12% in 2017 and 1.14% in 2016). The UAE population growth rate is above the world average fueled by immigrant influx looking for better employment and living opportunities, business.

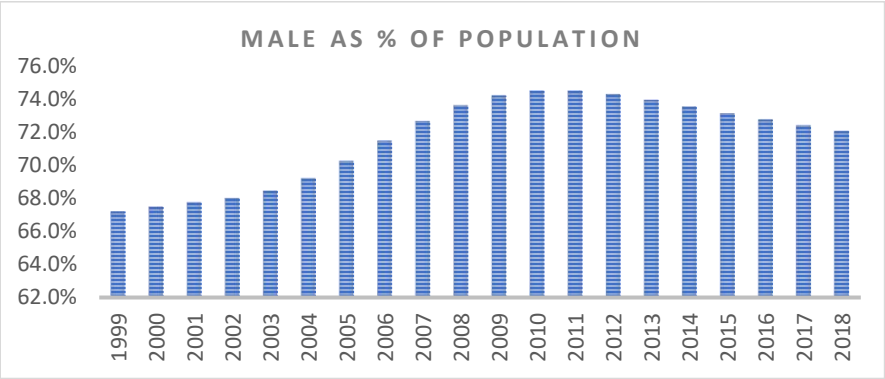


Graph 8: Total population counts all residents as per mid year values shared by world bank (world bank statistics 2018)

In the past 3 decades, population growth had been the lowest (0.7%) in the year 2014 but then it has been increasing and it reached to 1.5% in 2018. 5 years from 2005 to 2009 has seen growth rates more than 10% with an average rate of growth being 13.4%.

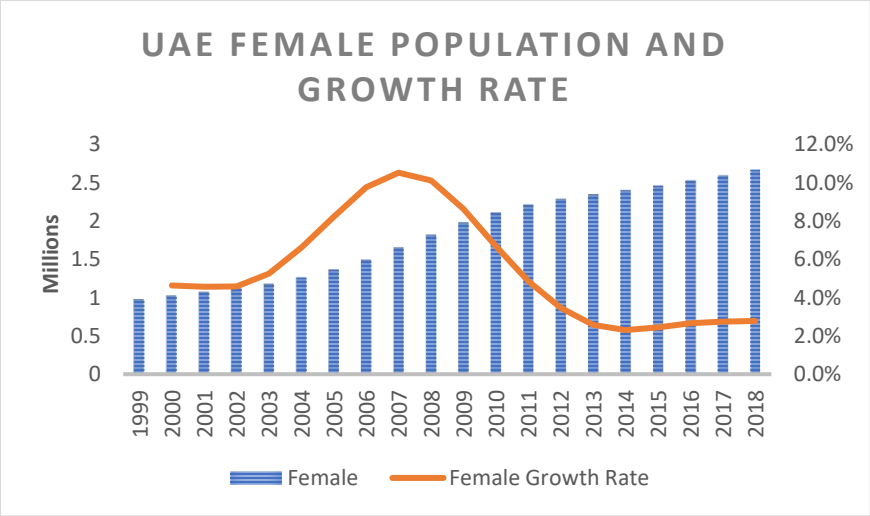
Population by Gender: *As per World Bank 2018* United Nations Population Division's World Population Prospects, males make up 72% of the UAE population, while the number of females accounts for only 28% of the total UAE population. This is due mostly to the high proportion of

men working in the construction industry and other similar labor class employment fueled by emigrantpopulation,



Graph9: Gender ratio statistics as per world banks stats

This gap was highest in 2011 as the male percentage reached 74.5%. Female contributes up 27.9% of the UAE population at 2.66 million. Ten-year female population growth rate has been 5.4% on an average from 2009 to 2018. For 2 years in 2007 and 2008 it had been more than 10%. Last 5 years average growth rate has been 2.6%.



Graph 10: Female population in the UAE and population growth statistics (World bank 2018)

1.3.1 Expats Vs. Emiratis Population statistics:

The UAE is home for expatriate from more than 200 countries. *World bank data* (International migrant stock % of population - United Arab Emirates) shows that the total percentage of foreigners in the is around 87.84% as of 2015. Exports have grown by 15% in the five years from 2005 to 2010. As of 2018 expats population stands at 88%.

Year	% of Expats in total population
1990	72%
1995	78%
2000	80%
2005	73%
2010	88%
2015	88%

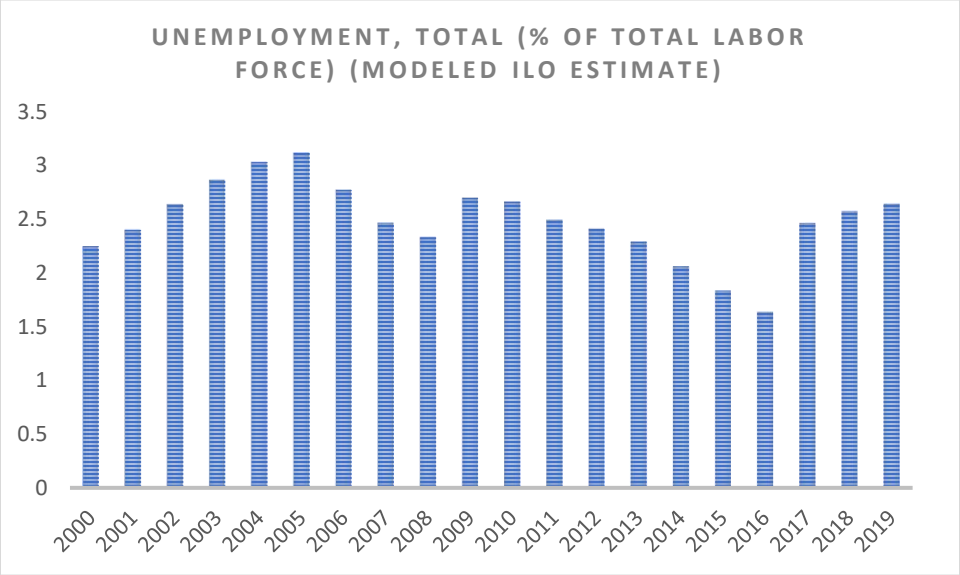
Table 10: International migrant stock (% of population) - United Arab Emirates (world bank statistics)

1.4 Unemployment:

The United Arab Emirates (UAE) have been an ILO member state since 1972. As per ILO estimates, despite a concerted push towards "Emiratization" of the workforce in the private sector, over 90 per cent of the private sector labor employees are expats. The UAE nationals continue to be employed in stable public sector jobs. Foreign workers in the UAE contribute to their home countries more than 29 billion US dollars in 2014, making the UAE the third biggest source of remittances in the world.

The unemployment rate for the last 20 years had been low and in the range of 1.6% to 3.1% as per ILO estimates. It continued to decline from a high of 3% in 2004 to 1.63% in 2016 but picked up in the last 3 years. The unemployment rate depicts the share of a country's labor force without

jobs, but available and actively seeking employment. The United Arab Emirates' unemployment rate is quite low in comparison to other gulf states.

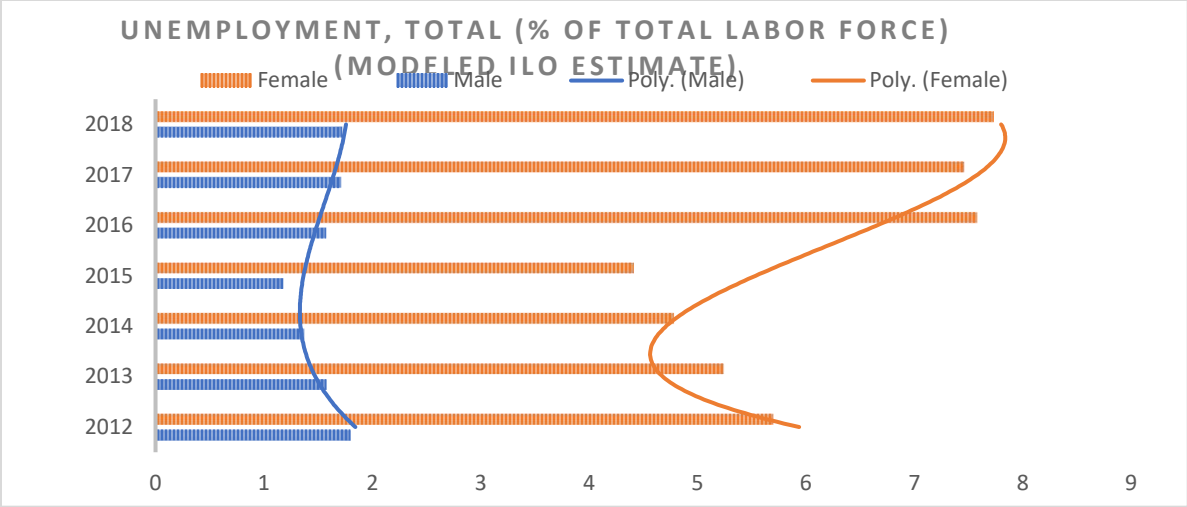


Graph 11: Unemployment, total (% of total labor force as per ILO estimates 2018)

The Low unemployment rate is credited to the fact that only working-age expatriates with jobs are allowed to stay in UAE.

Gender Unemployment statistics:

As per ILO statistics, female unemployment has been high of 6.09 percent on an average in last 10 years than male, with 1.7%. In 2017, the unemployment rate jumped from 4.4% to 7.5%. In the last 3 years, the female unemployment rate had been 7.59% on an average for female vs. 1.66% for the male.



Graph 12: Unemployment, total (% of total labor force) Male and Female (ILO estimates)

1.4.1 Unemployment rate by age group:

This statistic helps us understand Which age group has the highest rate of unemployment. As per UAE government (bayonet) statics, the youth unemployment rate for those between the ages of 16 and 24 was around 31% for male and 45% for female unemployment in 2017. Teenagers report the highest rates of unemployment.

Age group	Male	Female
15-19	27%	23%
20-24	4%	13%
25-29	2%	9%
30-34	1%	6%
35-39	1%	6%
40-44	1%	4%
45-49	1%	4%
50-54	1%	3%
55-59	1%	2%
60-64	1%	0%
+65	1%	0%

Table 11: Unemployment rate by age group and Gender (bayonet statistics UAE government 2017)

1.4.2 Labor force participation in UAE:

As per the Annual Economic Report 2018, The labor force participation rate was about 92.8 per cent of the total male population aged 15 years and above. The rate of participation in the female labor force was about 41.6 per cent of the total female population. Bayanat stats show that almost 62% of the workforce are primarily in the age group from 25 to 39 years. The employees are concentrated in the age 20-39 years, The percentage of employees in this group reached to 71.9% of the total employees.

Age Group	% of Labour workforce
< 16	0.0%
16 - 19	0.2%
20 - 24	9.9%
25 - 29	21.3%
30 - 34	23.2%
35 - 39	17.5%
40 - 44	11.3%
45 - 49	7.9%
50 - 54	4.5%
55 - 59	2.6%
60 - 64	1.2%
65 - 69	0.3%
70+	0.1%

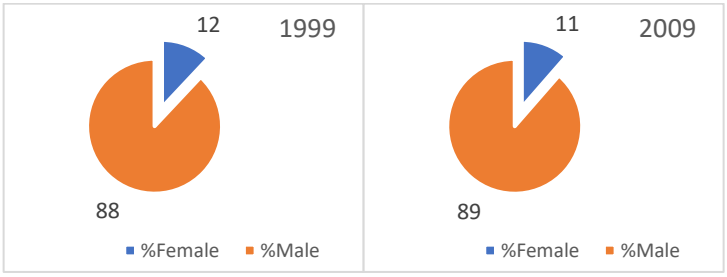
Table 12: Distribution of the workforce across age levels (as per Federal Competitiveness and Statistics Authority UAE)

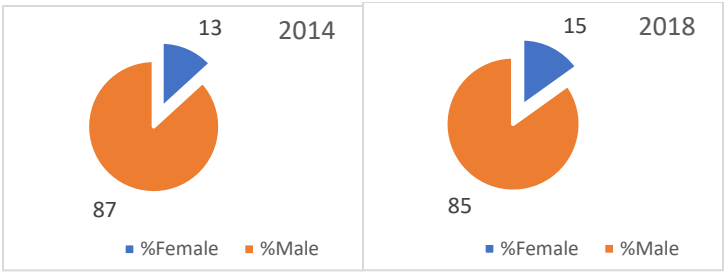
The employees are concentrated in the age 20-39 years, the percentage of employees in this group reached to 71.9% of the total employees This percentage is higher in females than males where the ratio was 89.0% among females while 89.5% among males. The percentage of the UAE's employees in this age group reached to 87.4%, where reaching to 94.3% among females while reaching to 83.2% among males.

Percentage Distribution of Employed Persons by Nationality,Gender and Age Group in United Arab Emirates, 2017									
Age Group	Emarati			Non Emarati			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
15-19	1.1%	0.6%	**0.9%**	0.5%	0.7%	**0.5%**	0.5%	0.7%	**0.5%**
20-24	12.4%	8.0%	**11.0%**	8.3%	9.6%	**8.5%**	8.4%	9.5%	**8.6%**
25-29	18.9%	22.2%	**19.9%**	18.1%	21.6%	**18.7%**	18.2%	21.6%	**18.8%**
30-34	19.8%	23.8%	**21.0%**	21.0%	22.8%	**21.3%**	21.0%	22.8%	**21.3%**
35-39	18.0%	20.9%	**18.9%**	17.4%	19.2%	**17.8%**	17.5%	19.3%	**17.8%**
40-44	12.9%	13.7%	**13.2%**	13.5%	12.8%	**13.4%**	13.5%	12.9%	**13.4%**
45-49	7.5%	7.3%	**7.4%**	9.6%	7.2%	**9.1%**	9.5%	7.2%	**9.1%**
50-54	5.0%	2.2%	**4.1%**	6.1%	3.1%	**5.5%**	6.0%	3.0%	**5.5%**
55-59	2.6%	1.1%	**2.1%**	3.6%	2.1%	**3.4%**	3.6%	2.0%	**3.3%**
60-64	1.2%	0.1%	**0.9%**	1.5%	0.6%	**1.3%**	1.5%	0.6%	**1.3%**
+65	0.7%	0.1%	**0.5%**	0.4%	0.3%	**0.4%**	0.5%	0.3%	**0.4%**

Table13: Labor force participation rate for ages and Gender (Federal Competitiveness and Statistics Authority UAE)

1.4.3 Female labor force participation: As per World Bank 2018 estimates based on age/sex distributions of the United Nations Population Division's World Population Prospects, males make up 85% of total labor force while the number of females accounts for only 15% of the total labor force.. Labor force comprises of 15+ and older who supply labor to produce goods and services. This has increased from 12% in 2010 to 15% in 2018.

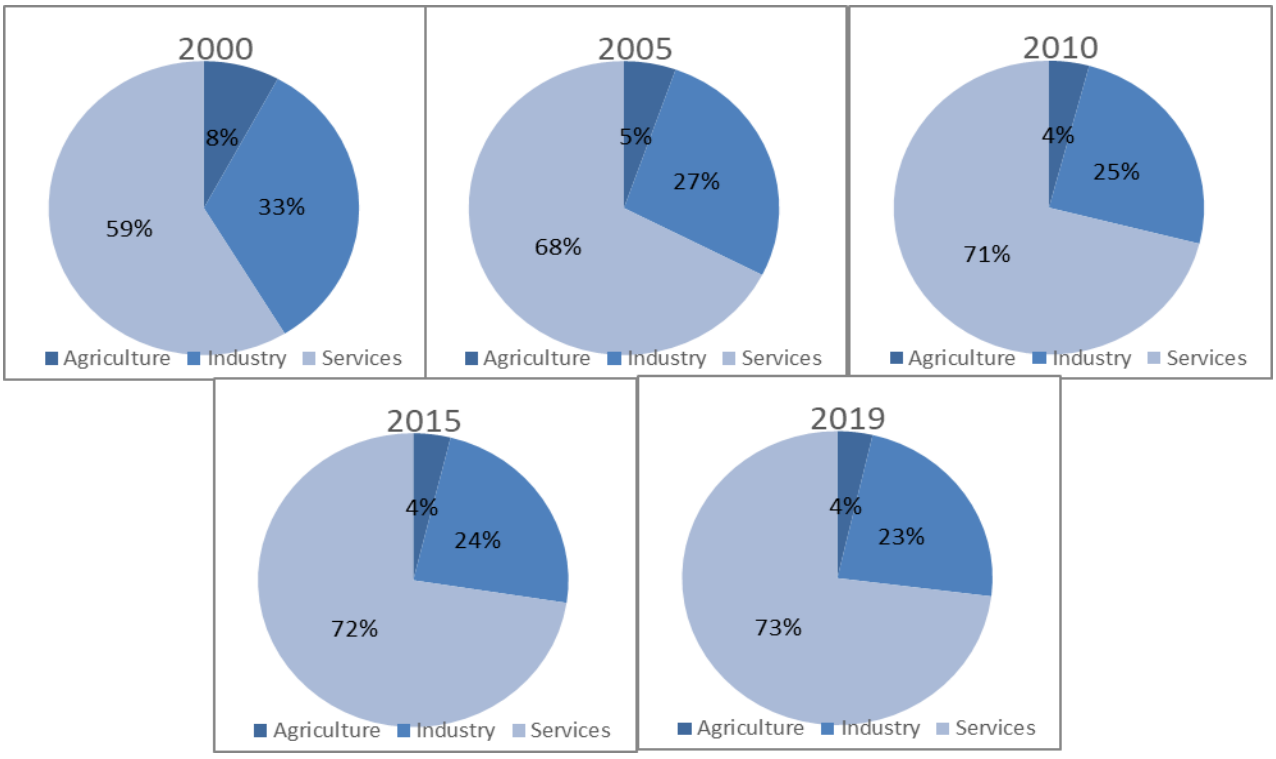




Graph 13: Female and Male as % of total labor force as per (World Bank statistics 2018.)https://datacatalog.worldbank.org/public-licenses#cc-by

1.4.4 Distribution of the workforce across economic sectors in the United Arab Emirates:

If we look at the distribution of the workforce across economic sectors, services sector contributing almost three fourth of the employment. The share of agriculture has fallen from 7.9% in the year 2000 to 3.7% in 2018. Industry share in the workforce has also shrunk by 10.1% in 19 years from 33.4% in 2000 to 23.3% in 2019.



Graph 14: workforce across economic sectors (bayanat. as statistics by UAE government)

1.4.4 Distribution of workforce by public and private sector: There is a clear divide in terms of preferences for Emiratis and non-Emiratis. Most (83.3%) of Emiratis are working with local and federal government where as expats in the private sector (75%). If we look at overall employment, almost 72% of the total workforce is employed in the private sector. Non profit organizations are mostly run by Emiratis supported by the law and the availability of capital.

Percentage Distribution of Employed Persons by Nationality,Gender and Sector in United Arab Emirates, 2017

Sector	Emarati			Non Emarati			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
Federal Government	34.9%	24.2%	**31.6%**	1.0%	0.7%	**1.0%**	2.3%	2.5%	**2.3%**
Local government	50.0%	55.4%	**51.7%**	6.5%	5.1%	**6.3%**	8.1%	8.9%	**8.3%**
Private Sector	7.4%	10.1%	**8.2%**	81.8%	43.9%	**75.0%**	79.0%	41.4%	**72.0%**
Shared (Government and Private)	6.9%	8.6%	**7.4%**	4.7%	4.0%	**4.5%**	4.7%	4.4%	**4.7%**
Foreign	0.2%	1.0%	**0.4%**	0.6%	1.0%	**0.7%**	0.6%	1.0%	**0.7%**
Diplomatic Authority	0.0%	0.0%	**0.0%**	0.1%	0.0%	**0.0%**	0.1%	0.0%	**0.0%**
Non-Profit Organizations	0.1%	0.3%	**0.2%**	0.0%	0.0%	**0.0%**	0.0%	0.0%	**0.0%**
Without Establishment	0.4%	0.2%	**0.3%**	0.1%	0.4%	**0.1%**	0.1%	0.3%	**0.1%**
Private Households	0.0%	0.1%	**0.0%**	4.6%	44.8%	**11.8%**	4.4%	41.4%	**11.3%**
Others	0.1%	0.0%	**0.0%**	0.0%	0.0%	**0.0%**	0.0%	0.0%	**0.0%**
Unknown	0.1%	0.1%	**0.1%**	0.6%	0.0%	**0.5%**	0.6%	0.0%	**0.5%**

Table 14: Percentage Distribution of employed persons by Nationality, Gender and Sector (as per Federal Competitiveness and Statistics Authority UAE)

1.4.5. Distribution of the workforce across job roles:

This statistic shows the relative distribution of the workforce across job roles in the UAE in 2017. In 2017, around 31.8% in the UAE were employed in crafts and related trades followed by 14.3 % in services sectors.

Managers	3.4%
Professionals	8.3%
Technicians and associate professionals	6.0%
Clerical support workers	8.8%
Service and sales workers	14.3%
Skilled agricultural, forestry and fishery workers	0.2%
Craft and related trades workers	31.8%
Plant and machine operators, and assemblers	10.0%
Elementary occupations	17.3%
Not Classified	0.0%

Table 15: Distribution of the workforce across job roles (as per Federal Competitiveness and Statistics Authority UAE)

Job Level	Male	Female
Managers	8.0%	6.5%
Professionals	14.0%	23.6%
Technicians and associate professionals	12.2%	9.5%
Clerical support workers	3.6%	7.7%
Service and sales workers	14.8%	10.0%
Skilled agricultural, forestry and fishery workers	1.0%	0.0%
Craft and related trades workers	22.8%	0.3%
Plant and machine operators, and assemblers	12.3%	0.5%
Elementary occupations	11.0%	41.8%
Not Classified	0.4%	0.2%

Table 16: Distribution of the workforce across job roles and Gender

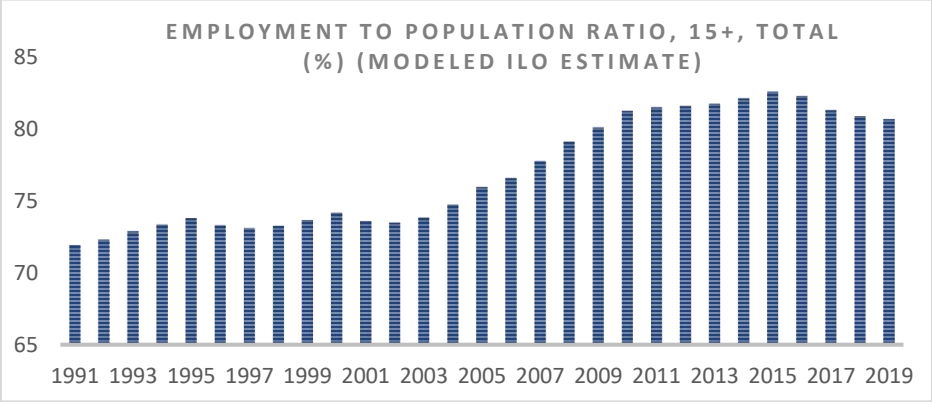
Percentage Distribution of employed persons by nationality, gender, and educational Level in United Arab Emirates: The percentage of the employees according to educational level varies considerably between Emiratis and non-Emiratis and between males and females. The educational level of expats is different from UAE employee. The percentage of expat employees, which have primary and less than primary 29.8%, but for Emiratis it is only 4.7%. The percentage of Emirati female employees who have got Primary and below education and below is 1%, but for expats it is 33.3%.

Educational level	Emarati			Non Emarati			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
Less Than Primary	1.5%	0.2%	**1.1%**	13.9%	20.5%	**15.1%**	13.4%	19.0%	**14.5%**
Primary	4.8%	0.8%	**3.6%**	15.1%	12.8%	**14.7%**	14.7%	11.9%	**14.2%**
Lower Secondary	14.8%	5.6%	**12.0%**	20.3%	10.4%	**18.5%**	20.1%	10.1%	**18.2%**
Upper Secondary	37.9%	27.5%	**34.7%**	18.1%	11.8%	**16.9%**	18.8%	12.9%	**17.7%**
Post-Secondary Non-Tertiary	6.9%	8.6%	**7.4%**	5.9%	5.2%	**5.8%**	6.0%	5.4%	**5.9%**
Bachelor or Equivelant	27.5%	49.7%	**34.3%**	20.8%	31.3%	**22.7%**	21.1%	32.7%	**23.3%**
Higher Education	6.6%	7.6%	**6.9%**	5.6%	7.8%	**6.0%**	5.6%	7.7%	**6.0%**
Adult Education	0.0%	0.0%	**0.0%**	0.2%	0.2%	**0.2%**	0.2%	0.1%	**0.2%**
Not classified	0.1%	0.0%	**0.0%**	0.0%	0.0%	**0.0%**	0.0%	0.0%	**0.0%**

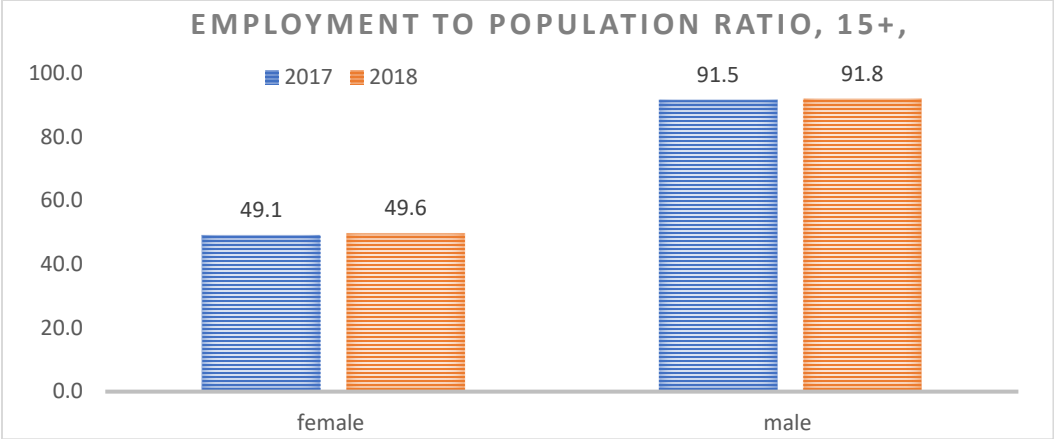
Table 17: Percentage distribution of Employed Persons by Nationality, Gender and Education (Federal Competitiveness and Statistics Authority UAE)

Employment to population ratio, 15+

The employment to population ratio is the estimation of the employed out of total population. Employment is defined as persons of working age that were engaged in any activity to produce goods or provide services for pay. Ages 15 and older are generally considered the working-age population. For female this ratio for UAE has improved from 49.6 in 2018 from 49.1% in from 2017.



Graph 15: Employment to population ratio, 15+ (world bank ILO estimates) for the UAE from 19991 to 2019



Graph 16: Employment to population ratio, 15+ (https://datacatalog.worldbank.org/public-licenses#cc-by)

Percentage distribution of unemployed by marital status and gender: As of 2017, The percentage of married Emiratis is 64.4% and for expats it is 73.8%. The percentage of Emirati's divorced employees reached 3.4% at the aggregate level where for non-Emiratis it is only 0.9%.

Marital status	Emarati			Non Emarati			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
Single	28.6%	38.6%	**31.7%**	23.2%	33.6%	**25.1%**	23.4%	34.0%	**25.3%**
Married	69.4%	53.2%	**64.4%**	76.5%	61.4%	**73.8%**	76.2%	60.8%	**73.4%**
Divorced	1.8%	6.8%	**3.4%**	0.3%	3.7%	**0.9%**	0.4%	4.0%	**1.0%**
Widowed	0.1%	1.5%	**0.5%**	0.0%	1.3%	**0.3%**	0.0%	1.3%	**0.3%**

Table 18: Percentage Distribution of Employed Persons by Nationality, Gender and Marital Status (as per Federal Competitiveness and Statistics Authority UAE)

Employed person by economic activity: A large percentage of the UAEs male population is working in the public construction industry, where the percentage of women workers is in the education sector.

Employed (Male) person by economic activity	Percentage
Construction	28.00%

Wholesale and retail trade & repair of motor vehicles	14.00%
Manufacturing	10.60%
Administrative and support service activities	7.00%
Transport and storage	6.90%
Accommodation and catering services	5.50%
Public administration and defense and compulsory social security	4.80%
Active households using individuals	4.60%
Professional, scientific and technical activities	3.60%
Financial activities and insurance activities	2.30%
Other	12.7%
Employed (Female) person by economic activity	**Percentage**
Active households using individuals	43.7%
Education	10.4%
Wholesale and retail trade & repair of motor vehicles	8.4%
Activities in the field of human health and social work	6.0%
Public administration and defense; and compulsory social security	4.3%
Administrative and support service activities	4.0%
Professional, scientific and technical activities	3.8%
Financial activities and insurance activities	3.3%
Accommodation and catering services	2.9%
Transport and storage	2.8%
Other	10.4%

Table 19: Percentage Distribution of Employed Persons by economic sector and Marital Status (as per Federal Competitiveness and Statistics Authority UAE)

1.5 Banking Sector:

The banking sector in the UAE is quite fragmented, with the market currently being served by 22 domestic banks, 7 banks from other Gulf countries and 31 foreign banks. The Central Bank of the

UAE is the primary financial regulatory authority. It is mandated to direct financial, credit and banking policy and oversee its execution in accordance with the state's general policy and in ways meant to help support the national economy and stability of the currency. The five big banks accounting for about 60% of the sector's total assets (CBUAE). Financial and Insurance activities contribution to GDP stands at 8.6%.

Number of Banks by Nationality of the Bank From 2011 - 2018								
Nationality of the Bank	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Local Bank	23	23	23	23	23	23	23	22
Gulf Bank	6	6	6	6	7	7	7	7
Foreign Bank	22	22	22	22	27	28	31	31
Total	51	51	51	51	57	58	61	60
Source: Central Bank Of The UAE								

Table 20: Number of Banks by Nationality of the Bank in UAE (as per Federal Competitiveness and Statistics Authority UAE)

1.5.1 Loans, advances, and overdrafts by loan:

The loan growth rate in the United Arab Emirates was 6.5% in 2018 over 2017. Corporate bank lending increased by an average of 15 between 2012 and 2018. Personal lending growth was 5% between 2012 and 2018, but -2.0% in 2017 and virtually no growth in 2018.

Loans, advances and overdrafts By Loans From 2011 - 2018								
Loans	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Real Estate Loans	240.8	253.842	287.3	277	288.4	317.9	350.4	379.1
Personal Loans	252.10	260.85	279.50	299.90	330.90	347.60	341.80	341.90
Bank Loans to Companies	391.40	395.03	462.70	521.10	593.20	646.30	703.40	761.20
Bank Loans to the Government	103.50	122.65	146.00	153.10	166.60	172.40	175.40	191.50
Total	987.80	1,032.38	1,175.50	1,251.10	1,379.10	1,484.20	1,571.00	1,673.70
Source: Central Bank Of The UAE(Billion AED)								

Table 21: Loans, advances, and overdrafts By Loans UAE (as per Federal Competitiveness and Statistics Authority UAE)

1.5.2 Net Interest from Banking Activities by nationality of the bank:

Almost 88% of net interest income of the banks is contributed by local banks.

Net Interest from Banking Activities By Nationality of the Bank From 2011 - 2018								
Nationality of the Bank	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Local Bank	39	40.26876	45	50.3	53.8	53.3	54.2	58.8
Gulf Bank	0.20	0.40	0.40	0.50	0.60	0.60	0.50	0.50
Foreign Bank	9.60	8.42	7.90	8.10	7.40	7.10	6.70	7.50
Total	48.80	49.09	53.30	58.90	61.80	61.00	61.40	66.80
Source: Central Bank Of The UAE (Billion AED)								

Table 22: Net Interest from Banking Activities by nationality of the bank in UAE (as per Federal Competitiveness and Statistics Authority UAE)

Total Liabilities to banks By Nationality of the Bank: Liabilities for local banks had grown by 8% in 2018 but for gulf and foreign banks it has gone down by 2%.

Total Liabilities for banks By Nationality of the Bank From 2011 - 2018								
Nationality of the Bank	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Local Bank	1316.2	1438.09	1630.7	1892.1	2088.8	2238.3	2333.4	2514.4
Gulf Bank	13.50	18.03	25.50	28.10	29.00	28.90	26.60	26.20
Foreign Bank	321.40	322.23	354.10	399.20	360.40	346.40	333.80	328.00
Total	1,651.10	1,778.35	2,010.30	2,319.40	2,478.20	2,613.60	2,693.80	2,868.60
Source: Central Bank Of The UAE(Billion AED)								

Table 23: Total Liabilities to banks By Nationality of the Bank in UAE (as per Federal Competitiveness and Statistics Authority UAE)

Deposits for banks By Nationality of the Bank: Deposits from local banks had grown by 10% in 2018 but for gulf it has gone down by 2% and foreign banks it has gone down by 5%.

Total Deposits for banks By Nationality of the Bank From 2011 - 2018								
Nationality of the Bank	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Local Bank	853.6	958.2	1057.4	1186.1	1250.4	1351.8	1419.8	1558.6
Gulf Bank	5.10	5.90	9.20	13.50	14.20	13.70	10.60	10.40
Foreign Bank	211.00	203.60	212.20	221.60	207.00	197.50	196.80	186.70
Total	1,069.70	1,167.70	1,278.80	1,421.20	1,471.60	1,563.00	1,627.20	1,755.70
Source: Central Bank Of The UAE(Billion AED)								

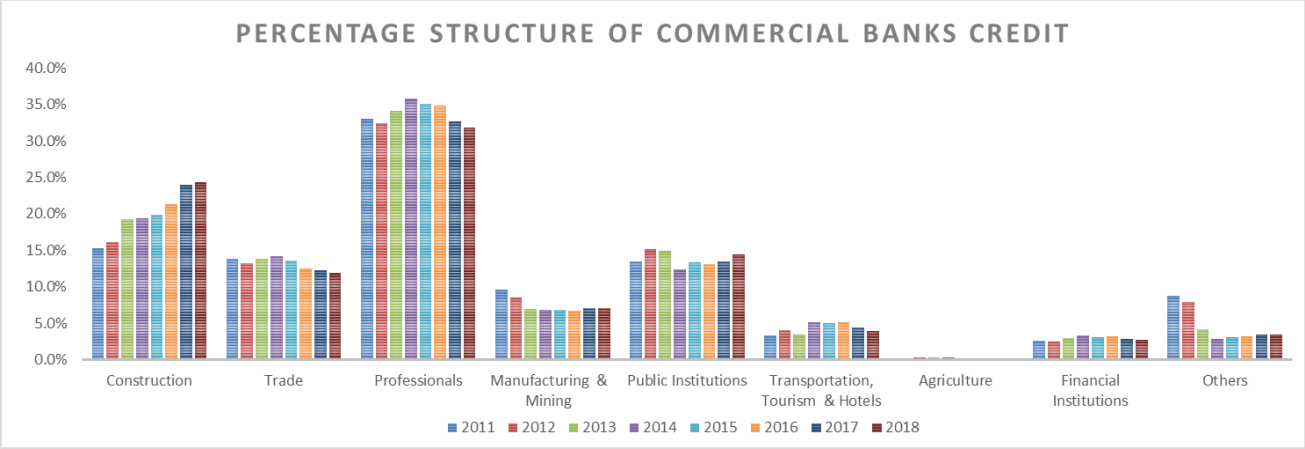
Table 24: Total Deposits from banks By Nationality of the Bank in UAE (as per Federal Competitiveness and Statistics Authority UAE)

1.5.3 Total Structure of Commercial Banks, Credit:

Growth in commercial credit has been primarily contributed by construction sector that has been increasing for the last 8 years.

Total Structure Of Commercial Banks Credit From 2011 - 2018								
Item	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Construction	426,253	474,443	664,506	780,588	863,077	972,040	1,094,868	1,158,366
Trade	387,154	390,067	477,516	572,035	593,803	571,104	561,043	565,736
Professionals	925,804	957,983	1,172,628	1,442,840	1,529,644	1,592,521	1,493,376	1,513,864
Manufacturing & Mining	268,916	250,193	239,605	272,942	297,596	304,825	320,740	336,601
Public Institutions	376,027	445,599	513,682	499,393	578,387	597,081	617,905	686,205
Transportation, Tourism & Hotels	93,142	119,606	118,440	207,474	217,253	233,019	201,742	189,118
Agriculture	4,485	7,547	10,055	9,913	5,732	5,044	7,809	7,283
Financial Institutions	72,905	73,534	100,620	131,824	135,600	144,329	127,498	129,782
Others	244,227	232,388	145,451	115,115	133,124	145,038	155,013	165,850
Source: Central Bank Of The UAE(Billion AED)								

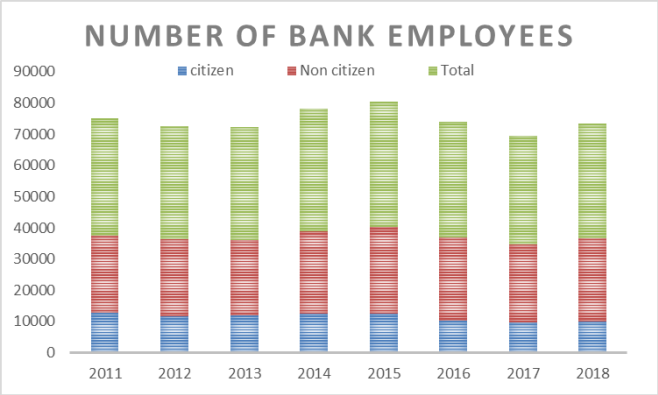
Table 25: Total Structure of Commercial Banks, Credit in UAE (as per Federal Competitiveness and Statistics Authority UAE)



Graph 17: percentage structure Of Commercial Banks, Credit in UAE (as per Federal Competitiveness and Statistics Authority UAE)

1.5.4 Number of Employees in Banking Sector:

UAE banks added 1954 employees in 2018, but banks reduced staff by almost 7.9% in 2016 compared to 2015. Nearly 27% of bakery workers are Emirati, but on average the share of Emiratis has been 31% over the past 8 years.



Graph 18: Number of Banking Employees in UAE (as per Federal Competitiveness and Statistics Authority UAE)

1.5.5 Number of Branches in Banking Sector:

Almost 88.6% of the total branches belong to the local banks. M&A and growth of digital banking have led to the 3% reduction in branches.

Number of branches From 2011 - 2018								
Nationality of the Bank	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Local Bank	768	828	864	892	897	869	793	782
Gulf Bank	7	8	10	10	10	11	11	11
Foreign Bank	76	105	105	102	102	100	106	90
Total	851	941	979	1,004	1,009	980	910	883
Source: Central Bank Of The UAE								

Table 26: Number of Branches in Banking Sector in UAE (as per Federal Competitiveness and Statistics Authority UAE)

1.5.6 Summary and Conclusion:

The UAE has a highly competitive index, high human development indicators, high ease of doing business, and an economy with a high per capita income and GDP. It also has a sizable annual trade surplus (2018). Successful efforts for economic diversification have reduced the share of oil in GDP and the non-oil sector's contribution to GDP has increased over time such as tourism and travel. The UAE economy is starting to recover from the 2015–16 slowdown caused by a decline in oil prices. The oil sector’s forecasts have also improved with higher oil prices in 2019

and predicted higher output. Inflation has been low, notwithstanding the introduction of the value-added tax (VAT) earlier in 2018.

The United Arab Emirates has a desert climate, characterized by pleasantly mild winters and very hot, sunny summers. Through there are more than 60 banks in the country, larger banks dominate the country's banking industry. Five biggest banks accounting for about 60% of the sector's assets. Banks remain liquid and decently capitalized. UAE banks – both domestic and foreign – are the backbone of the financial industry, providing retail and corporate services to members of the private and public sectors.

2. Chapter 2: Competency definition and studies from various Industries

The first step in learning how to define and explain competency is to clarify what it means. A competence is a skill or behavior that a person can perform. A competency can be described in several ways. The United Nations Industrial Development Organization defines it as "a set of knowledge, skills, and abilities that result in better performance." The College of Registered Nurses in British Columbia defines it as "the capacity to carry out measurable activities or outcomes". Other definitions include: *Aptitudes: An individual can be proficient in a certain area or set of tasks. A person can also demonstrate a competency without specialized knowledge. Another way to define competency is to look at it holistically. It can include generic skills, as well as attitudes and values. A person who can achieve a task requires a particular set of competencies. Similarly, a person must have certain leadership qualities to be effective in a leadership role. When an individual has the capability to perform a task, he or she is considered competent. In some cases, a person has multiple levels of competence, such as leadership or management.

2.1 Definitions of competency

The ever-shifting technological adoptions in organizations impact the labor market. Existing competences are becoming obsolete. The adoption of new practices leads to the creation of a new skill. Organizations need to assess the skills needed to do the job, identify the gaps, and develop a plan to fill the gaps. It thus becomes significant to develop the possibilities of identifying the employer demand for competencies and the supply of competencies (Lula at. al., 2019) R.E. Boyatzis (1982) finds that emotional, social, and cognitive intelligence competencies predict effectiveness in professional, management and leadership roles in many sectors of society. Woodruff (1991) proposed two pairs of competency concepts, first one refers to the areas of work in which the employee is competent and the second one is related to sets of behavior the employee must demonstrate to perform the professional tasks with competency. When it comes to identifying key competencies, there are various approaches suggested in literature. Variations are contributed mostly by the type of competencies as it could be

technological, physiological or behavioral. Some of the early work in this field has been done by McClelland (1973) where he determines competencies as any psychological or behavioral attributes connected with success. Competency refers to personal attributes that a person draws upon as part of their work activities, (Roberts (1997a). There is a wide range of competency definitions coming up in research publications, which varied in terms of constituent parts and competency elements and their acquirement and application as shown in Table1.

Author	Definition
R.E. Boyatzis (1982)	Competencies are the capacity of a given person to display behaviors compliant with the requirements of the job position specified by the organizational environment parameters, which, in turn, yields the desired results.
Drucker, 1985	Competence at the individual level as an ability of an employee to offer superior performance in assigned tasks
Spencer & Spencer, 1993	Provide superior performance in knowledge of assigned tasks, have authority to do something, highly skilled and awareness.
Barney, 1991;	Enabling the firm to implement strategies that exploit opportunities or neutralize threats in its environment' or 'make a disproportionate contribution to customer- perceived value
Foss & Knudsen, 1996	Company competencies are described as kinds of building blocks of company performance, and thus represent an internal-out view of companies
Boyatzis, Stubbs, &Taylor, 2002	Competence is an underlying characteristic of a person, motives, traits, abilities, aspects of the image or social role and knowledge that a person is able to use

Van Klink & Boon 2003	Competence is recognized as a bridge linking education and job requirements
Jackson & Schuler (2003)	Competencies are defined as 'the skills, knowledge, abilities, and other job characteristics that someone needs to perform a job effectively"
Lyle M. Spencer, Jr. & Signe M. Spencer (2004)	A competency is an underlying characteristic of an individual that is causally related to criterion that results in effective and/or superior performance in a job or situation
Atkociuniene, 2010	Competency is defined as as valuable, rare, non- replenishable and irreplaceable resources that can ensure competitive advantage for an organization in competitive environment

Table 1: Definition of competency

So, there are many studies on the subject literature to broadly classify them into behavior and professional segments.

2.2. Behavioral Competency

Behavioral Competency extremely valuable for assessing the quality of employees. Having a clear understanding of what competencies employees must possess is crucial for the development of an organization. Using behavioral competence measures can help managers make more informed decisions about which employees are right for the position. This can reduce hiring costs and lost productivity due to poor hiring decisions. This assessment also provides a clear picture of what skills are needed in the organization and in the individuals within it. In organizations, identifying behavioral competencies can be difficult. They are more difficult to measure than, say, sales numbers or performance goals. And because behavioral competencies are not as measurable as quantitative metrics, they are not always easy to implement. However, if implemented correctly, behavioral competency assessments can help organizations develop a more cohesive workforce. The assessment should include a clear understanding of what is expected of employees and how to promote this in the workplace. The competencies often relate

to job roles and are often assessed by an employer. Behavioral competence evaluations are often a part of the hiring process, and they can help an organization understand the needs of potential and current employees. In addition, employers often include these competencies in their job postings. When hiring, it's imperative to assess behavioral competencies as well as other attributes to make the selection process smoother and more productive.

In the first approach, competency is more behavior related and the based on the person's characteristics. According to Nusrat and Sultana (2019), competitive in the current marketplace and soft skills are “must-have” skills. Soft skill can be simply defined as abilities, attitudes, and behavior nature, rather than technical knowledge. Soft skills are considered as one of the vital abilities required in the business world. John (2009) mentions that modern corporations demand candidates who have the knowledge and soft skills that can be beneficial to the organization. Spencer and Spencer (1993) developed a competency dictionary consisting of 20 competencies distributed in 6 clusters. These generic competencies were illustrated with typical examples drawn from the interviews of superiors. . Companies are interested in four group of the core competencies such as intellectual, professional, personal and interpersonal (Cichoń M., Piotrowska I., 2018). The company’s strategic planning also has to be taken into account to make an analysis of labor demand (Fernández-Huerga E., 2019).

Oussii and Klibi (2017) stated that there is an association between communication skills and employment. Majid et. al. (2012) found that there was a correlation between students’ perception of the importance of soft skills towards employment and their study, he also hypothesized that teamwork skills would positively affect employment among graduates. Bandura (1997), Kanfer et al. (2001) and Van der Velde and Van den Berg (2003) discussed the significant role of self-efficacy in employability. Employers encourage the quality of purposeful thinking of graduates as this contributes to anticipate change (Harvey 1997). Lawrence (2006) discussed the importance of self-esteem in employability model. Self- esteem in educational psychology explains that an individual’s achievement is influenced by how he/she feels about

self. Boussiakou et al. (2006) considered emotional intelligence as benchmark in human resource management practices followed for recruitment and selection of job aspirants.

2.3. Professional competency

In simple terms, it is the habitual use of technical skills, communication, reflection, and clinical reasoning. It builds upon a foundation of scientific knowledge and basic clinical skills. It is a context-dependent, developmental construct. The definition of professional competence varies from person to person and from field to field. It is important for professionals to have a broad understanding of what constitutes professional competence and strive to continually improve it. Self-assessment is important throughout a professional career and should begin during training. The Roberts, Borden, Christiansen, and Lopez (2005) model may be used at different stages of a career. In fact, the American Psychological Association (APA) recognizes that continuing professional education is essential for maintaining and enhancing professional competence. Likewise, it is vital to evaluate one's own competence and to share that knowledge with others.

The ANF's standards for professional competence emphasize leadership. Although experienced practitioners often assume leadership roles, it is crucial to practice under a superior practitioner, who can provide guidance and inspiration. This is called collaborative leadership. Developing strong leadership skills is critical to advancing nursing practice. Using a mentor or a team is an excellent way to build a supportive, collaborative work environment. When you use your own strengths and resources to help others, you're also building a foundation for successful collaboration.

Another approach that focuses on performing professional duties so that organizations get the outcome and desired results (Whiddett, Hollyforde, 2003). From the employer’s perspective, they are concerned about finding suitable employees who no longer only have basic academic competencies, but better order thinking competencies like mastering, reasoning, thinking creatively, decision making and problem solving (Shafie and Nayan, 2010a). Carson and Carson (1998) discussed that emotional intelligence has a positive significant relationship with job experiences and emotional attachments to the job. There is also extensive work that has been done to find the gap between the existing competency and the needed competencies. Gallon et

al.(1995) suggested that identified competencies or capabilities are rated according to (a) the degree to which the capability has a direct impact on competitiveness, (b) the degree to which the capability constitutes best industry practice, and (c) the degree to which the capability has been optimized internally. Cichoń presented that there are 4 group of key competencies that are needed such as intellectual, professional, personal, and interpersonal that are most relevant for organizations (Cichoń M., Piotrowska I., 2018).

2.4. Match market demand with competency

The company's strategic planning is a key factor to decide the employees that need to be hired (Fernández-Huerga E., 2019). There has been significant work to match labor market demands with the competencies of graduates. Producing graduates with high employability skills—knowledge, attitudes, and the functional skill groups needed for acquiring and keeping a job (Heijde & Heijden, 2006). Harvey, Locke, and Morey (2002) defined employability as the ability to obtain, retain, and excel at a job. Heijde and Heijden (2006) defined employability as the use of an individual's competences (i.e., knowledge, skills, and abilities), to continuously acquire, fulfill, or create work. Weng and Tsai (2014) suggested an extended definition of employability as a set of knowledge, skills, and attributes that enable a person to choose a career and get employed, fulfill job duties, commit to work, evolve and advance in a career, recognize personal potentials, and feel satisfied and succeed.

2.5. Competencies from various industry sectors

Developing industry specific competency models is a complex process. A competency model should include the skills required for a job or position. These skills should be relevant and transferable, as they are often required in a specific industry. This is important, since competency statements can also be used to evaluate the strengths of individual employees. If you have a diverse staff, a competency statement will be useful in identifying and assessing the skills and knowledge needed for that role. A competency statement should also be relevant to the client's industry.

Myers et al. (2004) highlighted the supply chain manager's key competences such as decision-making skills, problem-solving skills, time management skills, social skills, and integrity among

over 20 relevant competences. Dischinger et al. (2006) explained a more complete and concise competence framework from six aspects of the supply chain: functional, technical, leadership, global management, and experience and credibility. Parikh (2014) in her study emphasized on the fact that competency mapping helps in managing human capital in an efficient manner. The paper explains certain management, functional, and behavioral competencies of sales managers in the retail industry. Chouhan (2013) attempts to develop a competency mapping model for HR professionals in the IT Industry for the purpose of the Training Need Assessment. Krishnaveni (2013) proposes to evaluate the capability of the employees of Meenakshi Mission Hospital and Research Centre, Madurai, India. It assesses numerous characteristics of employees' competency such as aptitude to develop a mutual association, communication, adaptability, leadership, and overall task proficiency. Madhavi (2019) attempted to find out the relevance of the interpersonal skills in the pharmaceutical industry were still the salespeople are considered as income producer. Andrew (2013) identified two predictive managerial competencies - composure and team leadership from the construction industry. Yuvaraj (2011) explained that the key job competencies required while working in a manufacturing industry are knowledge, ability and attitude.

2.6. Competencies in banking sector

Despite the importance of the customer experience, most banks are overly focused on IT transformation and business innovation. While open banking is growing in Europe, it lags in the Americas. Moreover, few banks are focused on customer journeys and innovative products and services. Instead, they are focusing on the core competencies of customer service, sales, and information technology. As a result, many banks are missing out on the most important aspects of customer experience. To address these challenges, ING created an on-demand talent pipeline that includes key competencies of future-oriented and digital banking. Despite the high-demand nature of the roles, the bank did not have an extensive infrastructure to develop its employees. The HR team quickly set up a process that matched between people with job responsibilities and prioritized the skills that were most in demand. The bank created a two-day training program for new hires that would prepare them for the challenges ahead. In the process, ING upgraded its

talent fluidity matching platform, a social media-like platform that connects individuals and teams across the globe. As a customer-centric bank, it is crucial to understand how and why customers use banking services. By using the right technology, banks can provide valuable insight to their customers and help them make better decisions. These tools can help banks develop a customer-centric approach to customer experiences and help them achieve their goals. However, if they lack these skills, it may be difficult to develop the appropriate solution for their needs.

There is limited research has been done in both non-financial and financial firms to map the employee competency and bank's performance. No studies have specifically dealt with individual effects of various employee competencies with organizational performance, particularly within banking organizations in the UAE.

Author	**Definition**
Papa, 1989	Communication competence
Payne, 2005	Communication competence
Ramo et al.,2009	Emotional and social competencies
Duc Duy Louis Nguyen(2015)	Age, education and experience
Chaudhary (2016)	Trait, Ability, Attitude,Skill, Knowledge
Salman(2020)	Social competence, team competence and communication competence

Table 2: Employee competency in Banking Sectors

At the decision level, institutions review the succession provisions to see how many individuals who have been employed fall into the high-potential category (McIlvaine, 1998). Flamholtz and Lacey's (1981) application of human capital theory focus directly on the skills of human beings in organizations. He identifies the importance of the distinct members of organizations as the important resource, rather than the practices and/or procedures used by the firm. Rashmi, K., &

Singh, R. (2020) build a model of employee engagement as it continues to be one of the top issues among organizations around the world as it is a key differentiator for organizational success. Jeffrey Pfeffer (1994) argues for the strategic role that staff management plays in creating new and long-term competitive advantage for the organization. Pfeffer observes that the fundamental idea behind competitive organizations are that if employees when well-managed they can provide the differentiating edge in competition. One of the challenges of current time is quickly evolving practices in the labor market. Adoption of new technology and practices demand the employees of the organization to develop new competencies at the faster rate. To remain viable in a global economy, it is critical that the businesses adapt to a fast-changing technological environment to meet their customer's needs. To stay competitive, organizations mean to evolve make all necessary changes to all, people, process, and technology to gain a competitive edge over their competition. The quality of a company's technology planning and implementation go a long way toward determining the success or failure of computerization and modernization (Hirschhorn, 1984). Employees sometime are used to the old practices and that may be the most beneficial situations for the organization (Caruth, Middlebrook and Rachel, 1985). Osei and Ackah (2015), in their study conducted in the pharmaceutical industry, explained that employee competencies have a significant and positive effect on organizational performance, and these immensely contribute towards achieving organizational objectives as well as its vision and mission.

2.7. Various competency models

There are many approaches to define employee competencies that differentiate competencies, including various typologies of competencies.

The literature and scientific studies (basic research) for competencies in banking sector in the UAE are even more sparse. The nearest subject literature that includes various typologies of competencies is suggested by Filipowicz (2014). He divided the competency schema into four groups, viz.

- Social – determining the quality performed tasks associated with contacts with people

(e.g., commercial contacts),

- Personal – related to the performance of tasks by the employee, and their job level affects the quality of the performed tasks,
- Managerial – involve human resource management, both with soft areas of management, work organization, as well as with strategic aspects of management,
- Professional (specialist, technical) – concern specialist tasks set for groups of positions.

Job competence assessment method—This is developed using interviews and observations of outstanding and average performers to determine the competencies that differentiate between them in critical incidents (Dubois, 1993).

Modified job competence assessment method—This also identifies such behavioral differences, but to reduce costs, interviewees provide a written account of critical incidents (Dubois, 1993).

Generic model overlay method—Organizations purchase an off-the-shelf generic competency model for a specific role or function (Dubois, 1993).

Customized generic model method—Organizations use a tentative list of competencies that are identified internally to aid in their selection of a generic model and then validate it with the input of outstanding and average performers (Dubois, 1993).

Flexible job competency model method—This seeks to identify the competencies that will be required to perform effectively under different conditions in the future (Dubois, 1993).

Systems method—This demand reflecting on not only what exemplary performers do now, or what they do overall, but also behaviors that may be important in the future (Linkage, Inc., 1997).

The competency schema- This concept was introduced in (Lula et al, 2019). It has been defined as a set competency and as a set of relations between them, together with the information about the importance of every competency and the importance of every relationship between any two competencies. We want to extend the Filipowicz (2014) and use the skills scheme defined by

(Lula et al, 2019) to build such for the banking sector in the UAE. The lack of connection between competencies is appeared in these observations of employability which means a set of achievements, understandings and personal qualities that increase the probability of getting a job and succeeding in a chosen profession (Cichoń M., Piotrowska I., 2018). Academic business cooperation is considered one of the instruments for bridging the skills gap between university graduates and business demand (Dávila C. et al, 2016).

3. Chapter 3: Text Mining: Methodological aspects

3.1 Introduction to Text Mining

Text mining is an Artificial Intelligence (AI) technology that uses Natural Language Processing (NLP) to extract meaningful information from inside large amounts of text data (Kehl 2020). NLP primarily comprises of natural language understanding (human to machine) and natural language generation (machine to human). The below figure depicts Text Mining, one of the Artificial Intelligence practices (Kumar, K. (2012))

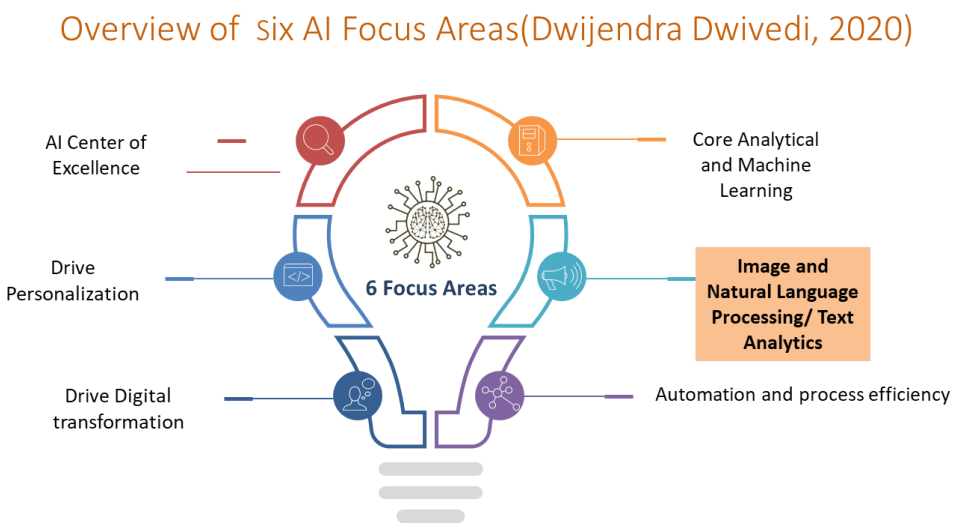


Figure 1: Text Mining, NLP as component of AI systems (Source: the author)

Text mining is the collection of methods to extract meanings, patterns, and structure unseen in unstructured textual data such text, audio, video, etc. (Chakraborty, Goutam 2013) It has a broader scope, and it is the process of discovering and extracting meaningful patterns and relationships from text collections. Text mining includes applications and algorithms for turning text into data and analyzing the data using statistical, visualization methods and natural language processing.

According to IDC (2021)80% of the data in an enterprise are unstructured (text, audio, video, etc.) And it is growing faster than structured data. As companies move towards data-driven culture, text mining is increasingly becoming a valuable resource. By analyzing the content on

websites and social media, companies can determine which topics and products are getting the most attention. A company can also analyze product reviews. This type of data can be useful for improving customer service and feedback. Regardless of the industry or field, text mining can help companies to make better decisions. Often, text mining tools have a number of applications. They can be used to extract complex information from text, such as a person's personality traits or communication preferences. Unlike traditional methods of data extraction, text mining tools can be highly effective in analyzing large volumes of data. These tools are especially useful for researching trends in a particular industry or the overall health of the population.

From the above illustration, we can see that Text Analytics using NLP are one of the focus areas of Artificial Intelligence. G. Miner (2012) has shown the intersection of text mining and six related fields, such as data mining, statistics, natural language processing, information extraction etc. This has been illustrated in figure 2.

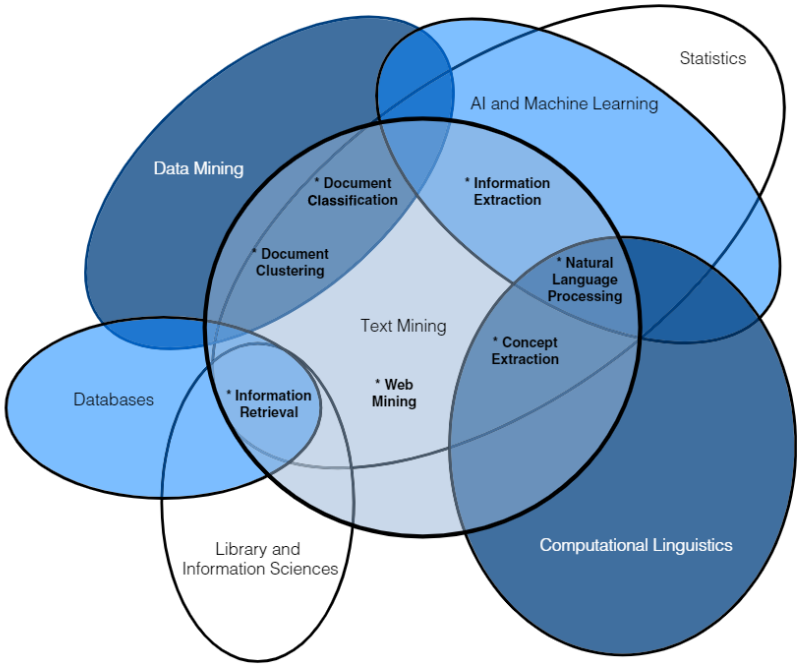


Figure 2: (G. Miner and others (2012)

The first computer application of summarization involved a large body of literature abstracts, Luhn (1958) used frequency of words and a measure of significance of important words along with distance of words from sentences to figure out sentence significance. The significant sentences were automatically selected to form abstracts. Doyle (1961) expanded Luhn's idea to suggest that document classification should use word frequencies as well as word associations. Claude Shannon's information theory (1941) is perhaps the most influential theory in information science. He showed that the digital interface between source and channel of the text was optimal because the creation of the text and the use of the text could be optimized by two separate coding problems with no loss of generality. Linguistics has developed two competing schools of thought: Chomsky's generative linguistics (rule-based syntactic structures) approach versus Luhn and Skinner's empiricist (deductive system) approach. Both are used in modern text analytics. Text analytics have evolved to include natural language processing techniques for extracting topics and summarizing content. To answer what methodology can be used for text analytics, we can use questions pathway created by G. Miner (2012) that includes 7 practice areas as explained below.

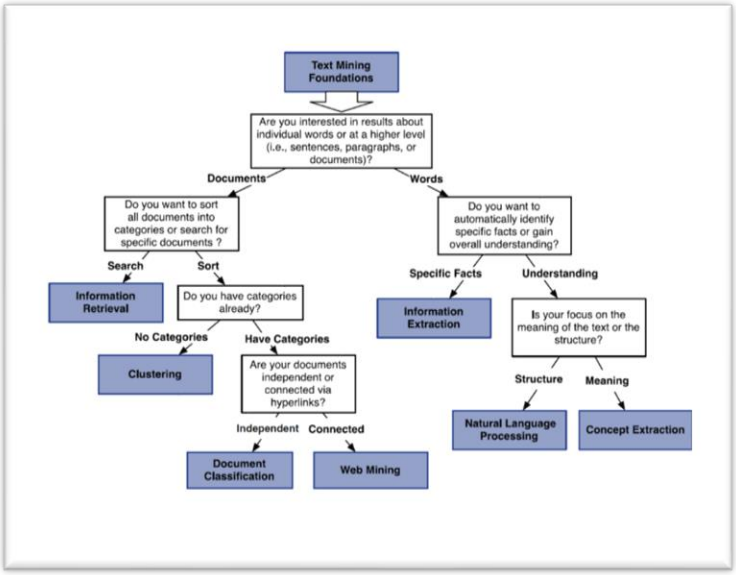


Figure 3: Project questions that can help determine the direction of text analytics (Miner, 2012)

3.2 Classification of problems considered in the text mining area

These seven practice areas are highly interconnected, and a typical text mining project will require methods from one or more areas based on applications:

1. Search and Information Retrieval (IR): IR is the method of searching for information in a document such as keyword search. (Mitra, Mandar et.al. 2000)
2. Document clustering: This creates mutually exclusive clusters to group and categorize documents or opinions or feedback text using various statistical segmentation and clustering methods. (Chakraborty, Goutam et.al. 2014)
3. Document classification: Classification of text and document based on supervised models. We use labeled data to train as a part of supervised modeling. (Pavlinek, Miha 2017)
4. Mining the Web: Data mining techniques are applied to web data to extract information about the behaviors of visitors to the web. (Eirinaki, Magdalini 2003)
5. Information Extraction (IE): In information extraction, relevant facts of interest areas are specified in advance. The goal is to automatically extract structured and relevant information from unstructured text. (Etzioni, Oren 2008)
6. Natural Language Processing is inspired from linguistically inspired technique. It is an area of Artificial Intelligence that helps computers understand languages. Natural language processing algorithms extract parts of speech from text. NLP uses a variety of methodologies to decode the ambiguities in human language Some of the steps are- Sentence Segmentation, Word Tokenization, Predicting Parts of Speech for Each Token, Text Lemmatization, Identifying Stop Words, Dependency Parsing Named Entity Recognition, Coreference Resolution etc.(Lauriola, Ivano 2022)
7. Concept extraction: concept extraction is an automated text analytics process that focuses on grouping of words into semantically corresponding sets. (Cohen, Aaron M. 2005)

3.3 Representation of documents in a form of frequency-matrix

The fundamental idea of applying classical data mining techniques to text mining relies on transforming text data (unstructured) to tables (or matrices) containing numbers (structured). Term Frequency and Inverse Data Frequency are used to measure the significance of a word in the data. It is especially useful for word notation in text-related computations, such as text analysis and NLPs.

3.3.1 Document term frequency:

It represents the number of documents that contain the term. DTM (Document Term Matrix) has been just the transpose of the Term Document Metrix. In a DTM, rows correspond to documents in the collection and columns correspond to terms. Salton (1963) published a visual depiction of a Document-Term Matrix. In this paper, he introduces the Document-Term Matrix by comparison to a kind of term-context matrix used to measure similarities between words. F.W. Lancaster (1964) published a comprehensive review of automated indexing and retrieval. Term frequency is the number of times a term occurs in a document. The rows represent the documents, and the columns correspond to the terms in the documents and the cells correspond to the weights of the terms. A term document matrix is a manner of representing words in text as an array (or matrix) of numbers. Term Document Matrix characterizes document vectors in matrix form where rows resemble to the terms in the document, columns resemble to the documents in the corpus and cells correspond to the weights of the terms. Weights could be binary or the frequency of the terms in the documents. Weights take the values- 0 or 1 where 1 represents the presence and 0 reflects the absence of the term. In the case of the term Frequency, the weights represent the frequency of the term in a specific document The basis specification for term frequency is as follows:

$$\mathbf{A} = \begin{bmatrix} f_{11} & \cdots & f_{1M} \\ \cdots & \cdots & \cdots \\ f_{N1} & \cdots & f_{NM} \end{bmatrix}$$

$$\text{Document Frequency(term t)} = \left(\frac{\text{Number of documents with the term t}}{\text{total number of documents}}\right) = \left(\frac{\text{d(t)}}{\text{n}}\right)$$

3.3.2 Inverse Term Frequency

Term Frequency (TF) and Inverse Document Frequency (IDF) were introduced by Spark Jones K. (1972) and contains two components: TF and ID. The measure of term specificity first proposed in that paper later became known as Inverse Document Frequency. As per Beel et. All (2015) TF-IDF is used by 83% of surveyed text-based research-paper recommender systems. Yates and Ribeiro-Neto (2011) proposed a term-weighting scheme that does not require access to the general document corpus and that considers information from the users' personal document collections. Van Rijsbergen (1979) and Manning et al. (2008) explained the role of IDF in information extraction.

The IDF is a measure of how much information the word provides, i.e., whether this is common or unusual in all documents. It is the logarithmically scaled inverse fraction of the documents that contain the word (obtained by dividing the total number of documents by the number of documents containing the term, and then taking the logarithm of that quotient). The IDF of a word calculates how unique or common a word is across a document collection. A unique word has a high IDF, while a common term has a low IDF. IDF for a specific word is calculated by dividing the number of documents (N) with the document frequency (DF) for a specific word in the document collection.

$$\text{Inverse Document Frequency} = \left(\frac{\text{Total number of documents}}{\textit{Number of documents with the term t}}\right) = \left(\frac{\text{n}}{\text{d(t)}}\right)$$

The logarithmic function is applied on the result to scale the quote for the length of the documents. Hence, normalizing the result and avoiding those words in long documents will obtain high IDF. Words in long documents tend to be repeated and consequently obtain high term frequency.

$$\text{IDF} = log\left(\frac{N}{d(t)}\right)$$

Words with a high TF–IDF weight are more significant than words with a lower TD–IDF weight. (Spärck Jones, K. 1972)

Then TD–IDF is calculated as

$$\text{tf–idf} = \text{Term Frequecy (t, d)} * \text{Inverse Dcoumement Frequecy (t)}$$

3.4 Identification of principal components using Singular Value Decomposition/ Latent Semantic Analytics – Algebraic approach

Topic extraction discovers the keywords in documents that capture the recurring theme of the text and is widely used to analyze large sets of documents to identify the most common topics in an easy and efficient way. LSA, also known as Latent Semantic Indexing or LSI, is a dimensionality reduction technique that typically operates on the term-by-document matrix by using a mathematical matrix decomposition technique called Singular Value Decomposition (SVD), which breaks down the original data into linearly independent components. LSA uses a bag of word template (BoW), which yields the term-document matrix (occurrence of terms in a document). The rows represent the words, and the columns represent the documents. LSA learns latent topics through a matrix decomposition on the document-term matrix using the singular value decomposition. LSA is generally used as a dimensional reduction or noise reduction technique. Another key advantage of LSA is a having common space for presentation of terms and documents and perform a low-rank approximation of the document - term matrix.

Deerwester et. al.(1990) took advantage of the implicit higher-order (or latent) structure in the association of terms and documents to reveal relationships. Foltz (1990) performed retrieval based on the “latent” semantic content of the documents rather than just on keyword matches. Gordon et. al.(1998) explored latent semantic indexing's effectiveness on two discovery processes: uncovering “nearby” relationships that are necessary to initiate the literature-based

discovery process; and discovering more distant relationships that may genuinely generate new discovery hypotheses. Berry et. al. (1995) found LSI a promising way to improve users access to many kinds of textual materials.

The steps for Latent Semantic Analysis are as follows:

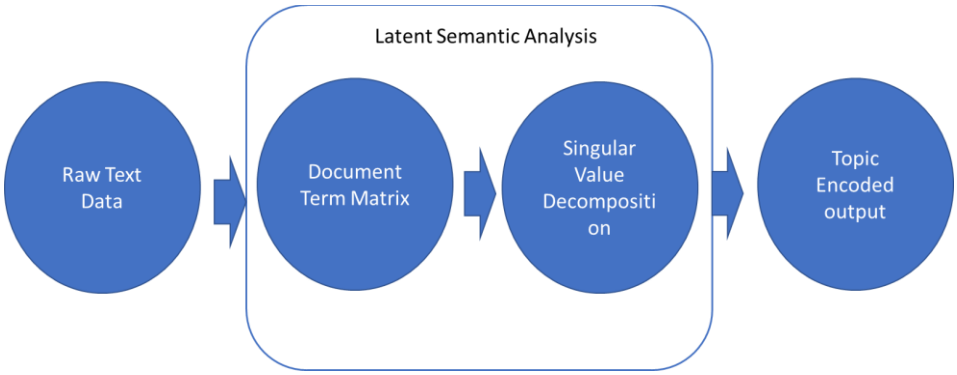


Figure 6 : Steps for Latent Semantic Analysis(source Autor)

3.4.1 Singular Value Decomposition (SVD)

Singular Value Decomposition (SVC) is a factorization method for matrices. It generalizes the eigen decomposition of orthonormal or square normal matrices and is related to polar decomposition. Its name derives from its properties, such as scalar product.

SVC is a common tool in statistical computing. It is a type of matrix transformation where vectors are transformed into an orthonormal basis. Singular Value Decomposition is an excellent mathematical tool for solving problems with complex matrices. It enables us to reduce a multidimensional matrix to its constituent parts and greatly simplifies matrix calculations. It was invented by differential geometers who wanted to find the equation of bilinear and square forms.

Given that if we have a term by document matrix A (m×n), it means there are m distinct terms in a document collection m≥n. The singular value decomposition of A is defined as

$$\mathbf{A} = \mathbf{USV}^T$$

Where m: number of text documents?

N: number of total unique terms (words)

K: number of topics that need to be extracted

U is the matrices of the term. A base using for presenting words

UV is the matrices of document vectors. ... Documents

S=diag(σ1,…,σn) is the diagonal matrix of singular values

Where $U^TU = I$ and $V^TV=I$ in a new common space and coordinates for

rows: US

columns: VS

3.4.2 Reduced Vector Space

For reducing the dimensions, we can simply choose the k largest singular values and the corresponding left and right singular vectors, and the best approximation of A with k-rank matrix is given by

$$\boldsymbol{A_K} = U_k \mathbf{S}\, V_k^T$$

U_k is comprises the first k columns of the matrix U

V_k is comprises the first k rows of matrix VT

diag(σ1,…σk) is the first k factors

The context-sensitive terms have higher similarity thus they will be close to each other in the new term space. This indicates that these terms are synonymous or multifocal. Then by re-combining the information in the initial feature space, LSA reduces vector space dimensions with most information retained. Similarities between the two documents and between term and document can be calculated in the new reduced vector space. The main advantage of SVD is that we can greatly reduce the matrix size from millions to 1000 or less.

3.4.3 An example of LSA application:

To see how this works, let's look at a small example from Deerwester et. al. (1990). He took A sample data set consisting of the titles of 9 technical memoranda. Terms occurring in more than one title are italicized. There are two classes of documents - five about human-computer interaction (c1-c5) and four about graphs (m1-m4). This dataset can be described by means of a term by document matrix where each cell entry indicates the frequency with which a term occurs in a document. The original matrix has nine columns, and 12 rows, each corresponding to a content word used in at least two of the titles as below:

Technical Memo Titles used here are:

c1: *Human* machine *interface* for ABC *computer* applications

c2: A *survey* of *user* opinion of *computer system response time*

c3: The *EPS user interface* management *system*

c4: *System* and *human system* engineering testing of *EPS*

c5: Relation of *user* perceived *response time* to error measurement

m1: The generation of random, binary, ordered *trees*

m2: The intersection *graph* of paths in *trees*

m3: *Graph minors* IV: Widths of *trees* and well-quasi-ordering

m4: *Graph minors*: A *survey*

Term by document matrix: 12 x 9 Type-by-Document Matrix With Type Frequencies Corresponding to the Titles as follows and represented as $[A]$=

Column1	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
computer	1	1	0	0	0	0	0	0	0

user	0	1	1	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
time	0	1	0	0	1	0	0	0	0
EPS	0	0	1	1	0	0	0	0	0
survey	0	1	0	0	0	0	0	0	1
trees	0	0	0	0	0	1	1	1	0
graph	0	0	0	0	0	0	1	1	1
minors	0	0	0	0	0	0	0	1	1

Table 1: term document Matrix

Where r (human.user) = -.38 r (human.minors) = -.29

The linear decomposition is shown as: $\mathbf{A} = \mathbf{USV}^T$ as in Table 2

{U} =

0.22	-0.11	0.29	-0.41	-0.11	-0.34	0.52	-0.06	-0.41
0.2	-0.07	0.14	-0.55	0.28	0.5	-0.07	-0.01	-0.11
0.24	0.04	-0.16	-0.59	-0.11	-0.25	-0.3	0.06	0.49
0.4	0.06	-0.34	0.1	0.33	0.38	0.0	0.0	0.01
0.64	-0.17	0.36	0.33	-0.16	-0.21	-0.17	0.03	0.27
0.27	0.11	-0.43	0.07	0.08	-0.17	0.28	-0.02	-0.05
0.27	0.11	-0.43	0.07	0.08	-0.17	0.28	-0.02	-0.05
0.30	-0.14	0.33	0.19	0.11	0.27	0.03	-0.02	-0.17
0.21	0.27	-0.18	-0.03	-0.54	0.08	-0.47	-0.04	-0.58
0.01	0.49	0.23	0.03	0.59	-0.39	-0.29	0.25	-0.23
0.04	0.62	0.22	0.0	-0.07	0.11	0.16	-0.68	0.23
0.03	0.45	0.14	-0.01	-0.3	0.28	0.34	0.68	0.18

{S} =

3.34								
	2.54							
		2.35						
			1.64					
				1.5				
					1.31			
						0.85		

							0.56	
								0.36

{V} =

0.20	0.61	0.46	0.54	0.28	0.0	0.01	0.02	0.08
-0.06	0.17	-0.13	-0.23	0.11	0.19	0.44	0.62	0.53
0.11	-0.5	0.21	0.57	-0.51	0.1	0.19	0.25	0.08
-0.95	-0.03	0.04	0.27	0.15	0.02	0.02	0.01	-0.03
0.05	-0.21	0.38	-0.21	0.33	0.39	0.35	0.15	-0.6
-0.08	-0.26	0.72	-0.37	0.03	-0.3	-0.21	0.0	0.36
0.18	-0.43	-0.24	0.26	0.67	-0.34	-0.15	0.25	0.04
-0.01	0.05	0.01	-0.02	-0.06	0.45	-0.76	0.45	-0.07
-0.06	0.24	0.02	-0.08	-0.26	-0.62	0.02	0.52	-0.45

Table 2: Complete SVD of matrix

Here we see a reconstruction based on just two dimensions that approximates the original matrix.. Each value in this new representation has been computed as a linear combination of values of the two retained dimensions, which in turn were computed as linear combinations of the original cell values. Note, therefore, that if we were to change the entry in any one cell of the original, the values in the reconstruction with reduced dimensions will be like below:

	c1	c2	c3	c4	c5	m1		m2	m3	m4
human	0.16	0.4	0.38	0.47	0.18	-0.05		-0.12	-0.16	-0.09
interface	0.14	0.37	0.33	0.4	0.16	-0.03		-0.07	-0.1	-0.04
computer	0.15	0.51	0.36	0.41	0.24	0.02		0.06	0.09	0.12
user	0.26	0.84	0.61	0.7	0.39	0.03		0.08	0.12	0.19
system	0.45	1.23	1.05	1.27	0.56	-0.07		-0.15	-0.21	-0.05
response	0.16	0.58	0.38	0.42	0.28	0.06		0.13	0.19	0.22
time	0.16	0.58	0.38	0.42	0.28	0.06		0.13	0.19	0.22
EPS	0.22	0.55	0.51	0.63	0.24	-0.07		-0.14	-0.20	-0.11
survey	0.10	0.53	0.23	0.21	0.27	0.14		0.31	0.44	0.42
trees	-0.06	0.23	-0.14	-0.27	0.14	0.24		0.55	0.77	0.66

graph	-0.06	0.34	-0.15	-0.30	0.20	0.31		0.69	0.98	0.85
minors	-0.04	0.25	-0.10	-0.21	0.15	0.22		0.50	0.71	0.62

r (human.user) = .94 r (human.minors) = -.83

Table 3: term document Matrix

reconstructed two-dimensional approximation

In the original table, human never appears in the same passage with either user or minors —they have no co-occurrences, contiguities or "associations" as often construed. The correlations (using Spearman r to facilitate familiar interpretation) are -0..38 between human and user, and a slightly higher -0.29 between human and minors . However, in the reconstructed two-dimensional approximation, because of their indirect relations, both have been greatly altered: the human-user correlation has gone up to 0.94 and the human-minors correlation up to -0.83. Thus, because the terms human and user occur in contexts of similar meaning, even though never in the same passage, the reduced dimension solution represents them as more similar, while the opposite is true of human and minors.

Deerwester et. al.(1990) used the two-dimensional plot of 12 Term and 9 Documents from a sample set. Terms are represented by filling circles. Documents are shown as open squares, and component terms are indicated parenthetically. The query ("human computer interaction") is represented as a pseudo-document. Axes are scaled for Document-Document or Term-Term comparisons. The dotted cone (from figure 8) represents the region whose points are. Within a cosine of .9 from the query 4. All documents about human-computer (clLc5) are "near" the query (i.e., within this cone), but none of the graph theory documents (ml-m4) arc nearby, In this reduced space, even documents c3 and c5 which share no terms with the query arc near it.

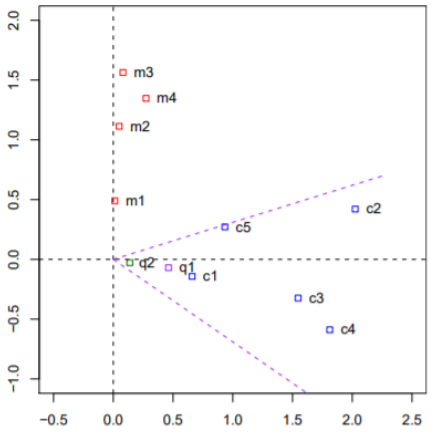


Figure 8 Ref: Deerwester, Scott, et al. R1 representation of the c1–c5 and m1–m4 documents (1990).

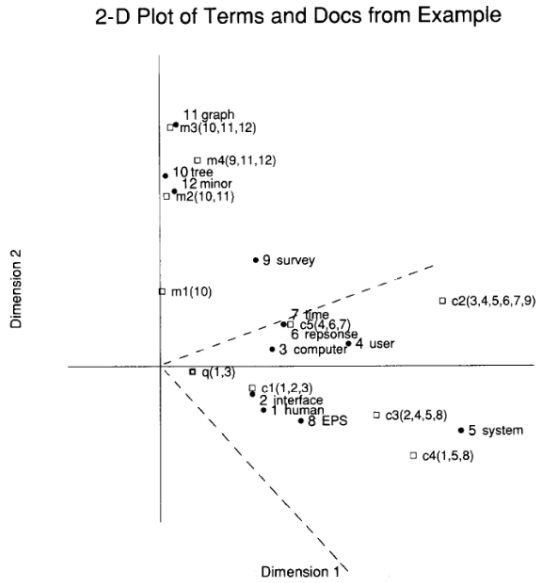


Figure 9 Ref: Deerwester et. al.(1990) "Indexing by latent semantic analysis."

3.5 T-distributed Stochastic Neighbor Embedding

T-SNE is an unsupervised machine learning algorithm that transforms high-dimensional data into low-dimensional ones using a nonlinear metric called dimensionality reduction. It can be applied to data in which the distances between two variables are small, as in a single-cell analysis.

However, it is not without drawbacks. In fact, T-SNE is largely underutilized, as it is only used to embed a single-cell set of gene expression measurements.t-SNE is a statistical approach to visualization high-dimensional data. Its goal is to find a low-dimensional representation of each data point, a technique known as'resizable encoding'. The t-SNE algorithm is based on the original Stochastic Neighbor Embedded (SNE) method, developed by Geoffrey Hinton and Sam Roweis. Laurens van der Maaten proposed an improved version of this method, called t-SNE. The technique models, each high-dimensional object with a two-dimensional point, where the number of points is high and the distance is low. T-SNE solves this problem by mapping high-dimensional data in two or three dimensions. The number of degrees of freedom in t-SNE varies according to the dimensionality of the data. An embedding in thirty dimensions is less likely to produce a low-dimensional visualization than a dataset in two dimensions. This difference is one of the main advantages of t-SNE.

It is a nonlinear dimensionality reduction technique well-suited for embedding high-dimensional data for visualization in a low-dimensional space of two or three dimensions. tSNE, (t-distributed stochastic neighbor embedding) is a clustering technique that has a similar end result to Principal Component Analysis (PCA). It helps in understanding high-dimensional data and project it into low-dimensional space. Laurens van der Maaten et. al. (2011) applied multiple maps t-SNE to a large data set of word association data and to a data set of NIPS co-authorships, demonstrating its ability to successfully visualize non-metric similarities. Moafer(2014) investigated the acceleration of t-SNE—an embedding technique that is commonly used for the visualization of high-dimensional data in scatter plots.

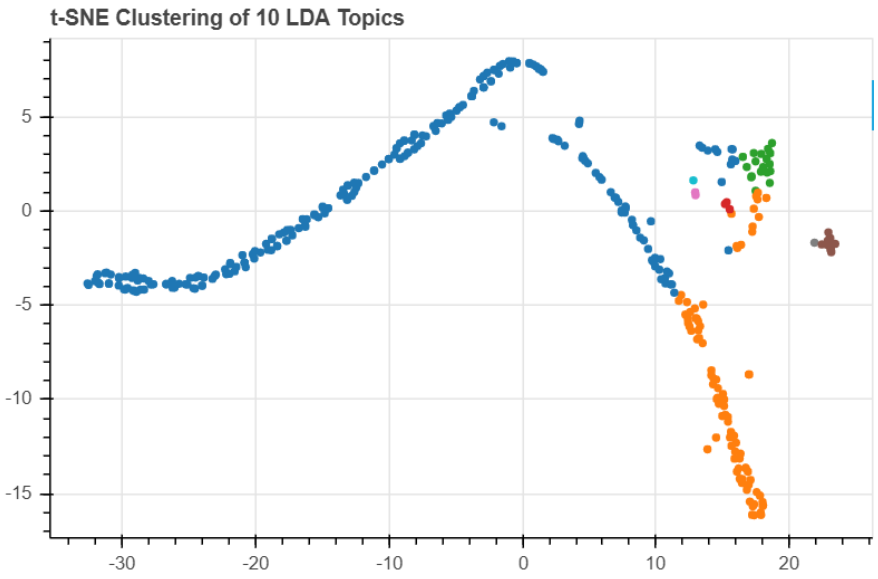


Figure 10: T-SNE plots (Source- Author)

3.6 Topic Modeling using LDA based probabilistic model:

3.6.1 Informal Introduction:

LDA is a probabilistic generative model that makes it possible to explain a collection of discrete observations, such as sentences, by unobserved (latent) subjects. LDA envisages a fixed set of topics. Each topic represents a group of words. The goal of the LDA is to identify topics in documents in a way such that topics are described as a mixture of words and the contribution of every word in this mixture is different. Then each document may be presented as a mixture of topics that have been previously identified (also the contribution of each topic in the given document is varied). The purpose of the LDA is to map all documents to topics in a manner, so that the words in each document are primarily captured by these imaginary subjects. In modeling LDA topics, we create numerous groups of different topics. However, we do not know how many groups there are. So, there are going to be different types of groups. Next, we look at and compare topic modelling, and decide which topic model is the most logical, the most significant, have the clear distinction in the model. Next, the group (model) that makes the most

sense will be chosen from all groups of topics. LDA does not expressly label classes. The onus is on the user to understand the class labels.

3.6.2 References

Blei. et. al.(2012) described LDA as a generative probabilistic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics. LDA is a model closely linked to the probabilistic latent semantic analysis (PLSA) by Hofmann (1999), an application of the latent aspect method to the latent semantic analysis task. D. Blei (2002) explained how LDA extends PLSA method by defining a complete generative model. Rosen-Zvi et al. (2004) paper illustrated that the topic proportions are attached to authors, and it allowed for inferences about authors, for example, author similarity. Blei and McAuliffe(2007) came up with a general-purpose method for incorporating metadata into topic models. Gottipati et al. (2018) leveraged topic modeling and data visualization methods to analyze student feedback comments from seven undergraduate courses. Al-Obeidat et al. (2018) further proposed an opinions sandbox for topic extraction, sentiment analysis for pulling issues and their associated sentiments from a database. They used LDA for topic extraction and the "bag-of-words" sentiment analysis algorithm, where polarity is determined based on the frequency of positive/negative words in a document. Asmossen & Moller (2019) presented a framework to leverage the topic modeling technique for performing an exploratory literature review of an extensive collection of papers. Korzycki et. al. (2017) posits that LDA treats documents as probabilistic distribution sets of words or topics. These topics are not strongly defined – as they are identified on the basis of the likelihood of co-occurrences of words contained in them." These models are generally good at grouping words together into topics. Topic modeling is not mutually exclusive, and a document can be allocated to multiple topics.

Graph-based models in text mining represent co-occurrence of words in text segments or in documents. They are useful for identification of keywords and key-phrases. Stuart Rose (2010) proposed the rapid, automatic keyword extraction (RAKE), an unsupervised, domain-independent, and language-independent method for extracting keywords from individual documents. Mihalcea and Tarau (2004) described another system where co-occurrences of the selected words within a fixed-size sliding window are accumulated within a word co-occurrence graph. A graph-based ranking algorithm (TextRank) is applied to rank words based on their associations in the graph, and then top-ranking words are selected as keywords. Keywords that are adjacent in the document are combined to form multi-word keywords. An ontology represents knowledge as a set of concepts within a domain, and the relationships between those concepts. It can be used to reason about the entities within that domain and may be used to describe the domain. An ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them. Gruber (1993) defines an ontology as "a formal, explicit specification of a shared conceptualization." Common goals in developing ontologies are to share a common understanding of the construction of information among people (Musen 1992; Gruber 1993). Ontologies expressed using the Ontology Web Language (OWL) can provide context representation, meta-language definitions, and semantic reasoning capabilities (Thomas et al., 2005). The Ontology Web Language (OWL) has evolved as the popular open standard for semantic knowledge representation and was developed as part of the Semantic Web initiatives from W3C (Horridge et al., 2004; Horrocks, et al., 2008). Ontology development is an iterative process and requires the researcher to gain a thorough understanding of the problem domain. It is common practice to verify the suitability of existing ontologies by testing against well-defined use cases (Chen, 2004; Firat et al., 2004). Available ontologies can be extended, or new ontologies created. Ontology-based identification of fragments of the offer can be implemented on different platforms, including the Web. It is a useful way to search large sets of ontology terms, and many studies have used it in recent years. In a recent study, Robles-Bykbaev and colleagues described the use of a formal knowledge base to support data analysis and inference processes.

3.6.3 Dirichlet Distribution: Dir($\boldsymbol{\alpha}$)

This distribution is often denoted as Dir($\boldsymbol{\alpha}$). The Dirichlet distribution is a continuous multivariate probability distribution, which is often abbreviated as Dir. It is parameterized by a vector of positive reals, and is known for its smoothness. This type of distribution is characterized by its large central limit and low tail. The Dirichlet distribution is often used as a prior distribution in Bayesian statistics. It is a conjugate distribution for category and multinomial distributions. The symmetric Dirichlet is a special case of the Dirichlet distribution, with the same value in all elements. It is often used as a vague prior distribution, with the concentration parameter as a constant. The alpha value is called the concentration parameter, and it is inversely related to the variance around the mean. The Dirichlet distribution is a multivariate probability distribution with no positive numbers. Its parameters do not have to be integers, but must be positive real numbers. The parameters do not need to be normalized or divided. The Dirichlet distribution has a skew distribution, and it is often used to describe the likelihood of a specific event. It is also used in Bayesian statistics. It is named after Johann-Dietrich-Drichlet, a Belgian mathematician. The length of the vector $\boldsymbol{\alpha}$ denotes the dimension of the Dirichlet. LDA uses Dirichlet to put a prior on the topic distribution. It is often used as prior distributions in Bayesian statistics.

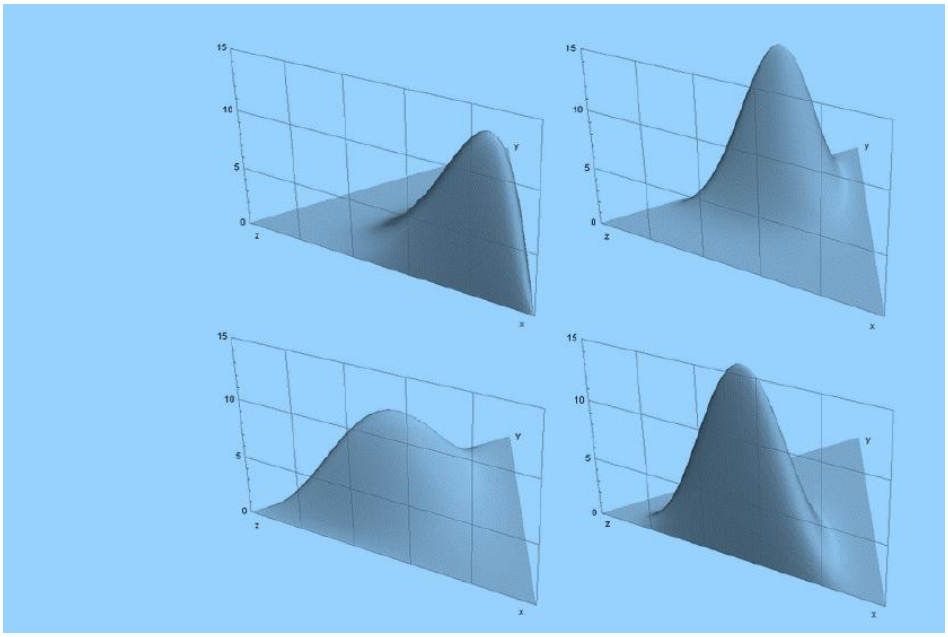


Fig 11: Dirichlet Distribution examples from wiki

Dirichlet distribution is an exponential family distribution over the simplex of positive vectors that sum to one.

- A word is the basic unit of discrete data, defined to be an item from a vocabulary indexed by {1, . . . ,V }. We represent words using unit-basis vectors that have a single component equal to one and all other components equal to zero. Thus, using superscripts to denote components, the vth word in the vocabulary is represented by a V -vector w such that wv = 1 and wu = 0 for u = v.
- A document is a sequence of N words denoted by w = (w1, w2, . . . , wN), where wn is the nth word in the sequence.
- A corpus is a collection of M documents denoted by D = {w1, w2, . . . , wM}.

LDA assumes the following generative process for each document w in a corpus D:

- 1. Choose N ∼ Poisson(ξ).
- 2. Choose θ ∼ Dir(α).
- 3. For each of the N words wn:

(a) Choose a topic zn ∼ Multinomial(θ).

(b) Choose a word wn from p(wn | zn, β), a multinomial probability conditioned on the topic zn.

A k dimensional Dirichlet random variable has the following probability density on this simplex: Density of Dirichlet is given by:

$$p(\boldsymbol{\theta}|\boldsymbol{\alpha}) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k \theta_k^{\alpha_k - 1}$$

1. Per-document topics proportions θ_d is a multinomial distribution, which is generated from Dirichlet distribution parameterized by α. (For each document, d)
2. Similarly, topics β_k is also a multinomial distribution, which is generated from Dirichlet distribution parameterized by η.
3. For each word n, its topic $Z_{d,n}$ is drawn from document topic proportions θ_d.
4. Then, we draw the word $W_{d,n}$ from the topic β_k, where $k = Z_{d,n}$.

3.6.4 LDA as generative model:

LDA is a generative model. A generative model describes how a dataset is generated, in terms of a probabilistic model. By sampling from this model, we can generate new data. Approaches that explicitly or implicitly model the distribution of inputs as well as outputs are known as generative models.

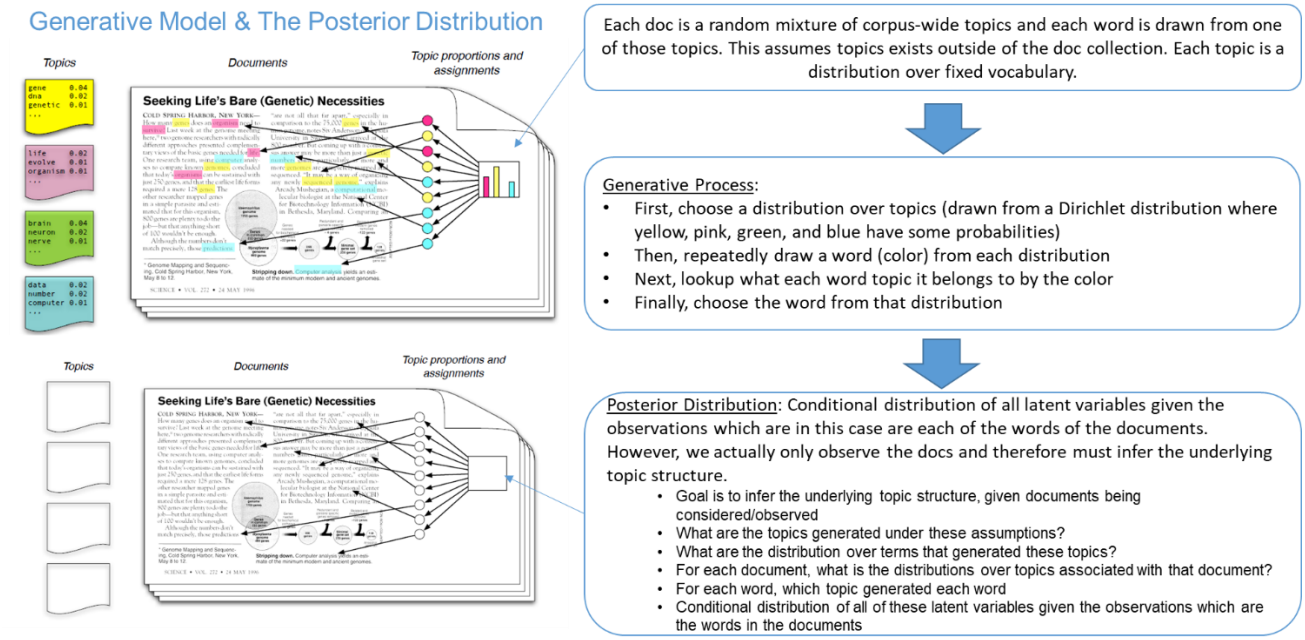


Fig 10: Generative model for latent Dirichlet allocation(Source Bobby 2015)

It is necessary to underline the meaning of these two matrices:

phi (ϕ) : Is the word distribution of each topic, i.e. The probability of each word in the vocabulary being generated if a given topic, z (z ranges from 1 to k), is selected.

$$\emptyset = \begin{bmatrix} \emptyset_{11} & \dots & \emptyset_{1,LV} \\ \dots & \dots & \dots \\ \emptyset_{LD,1} & \dots & \emptyset_{LT,LV} \end{bmatrix}$$

theta (θ)) : Is the topic proportion of a given document. More importantly, it will be used as the parameter for the multinomial distribution used to identify the topic of the next word. To clarify, the selected topic's word distribution will then be used to select a word w.

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_{11} & \dots & \theta_{1,LV} \\ \dots & \dots & \dots \\ \theta_{LD,1} & \dots & \theta_{LT,LV} \end{bmatrix}$$

We also need to introduce three other parameters mainly:

alpha (α) : To determine the value of θ , the topic distribution of the document, we sample from a Dirichlet distribution using α as the input parameter. For example: creating a document generator to mimic other documents that have topics labeled for each word in the doc. I can use the total number of words from each topic across all documents as the β values. Alpha parameter is Dirichlet prior concentration parameter that represents document-topic density with a higher alpha, documents are assumed to be made up of more topics and result in more specific topic distribution per document. Low α means A document is more likely to be represented by just few of the topic.

beta (β) : Beta parameter is the same prior concentration parameter that represents topic-word density — with high beta, topics are assumed to made of up most of the words and result in a more specific word distribution per topic. Low β means topic may contain a mixture of just a few of words. In order to determine the value of ϕ , the word distribution of a given topic, we sample from a Dirichlet distribution using β as the input parameter. β values are our prior information about the word distribution in a topic. For example: I am creating a document generator to mimic other documents that have topics labeled for each word in the doc. I can use the number of times each word was used for a given topic as the β values.

xi (ξ) : In the case of a variable length document, the document length is determined by sampling from a Poisson distribution with an average length of ξ .

also

k : Topic index

z : Topic selected for the next word to be generated.

w : Generated Word

d : Current Document

Logic for generative model could be represented as below:

---

1. For k = 1 to K where K is the total number of topics
    - *Sample parameters for word distribution of each topic*
    - $\phi^{(k)} \sim$ Dirichlet(β)
2. For d = 1 to D where number of documents is D
    - *Sample parameters for document topic distribution*
    - $\theta_d \sim$ Dirichlet(α)
    - For w = 1 to W where W is the number of words in document *d*
        - *Select the topic for word w*
        - $z_i \sim$ Multinomial(θ_d)
        - *Select word based on topic z's word distribution*
        - $w_i \sim$ Multinomial($\phi^{(zi)}$)

Fig 11 . Generative model for LDA

The three steps could be explained as below.

3.6.4.1 Learning: We treat data as observations that arise from a generative probabilistic process that includes hidden variables % document generation process.

3.6.4.2 Updating the variables according to the observations: Infer the hidden structure using posterior inference.

3.6.4.3 Inference using learnt model: Situate new data into the estimated model.

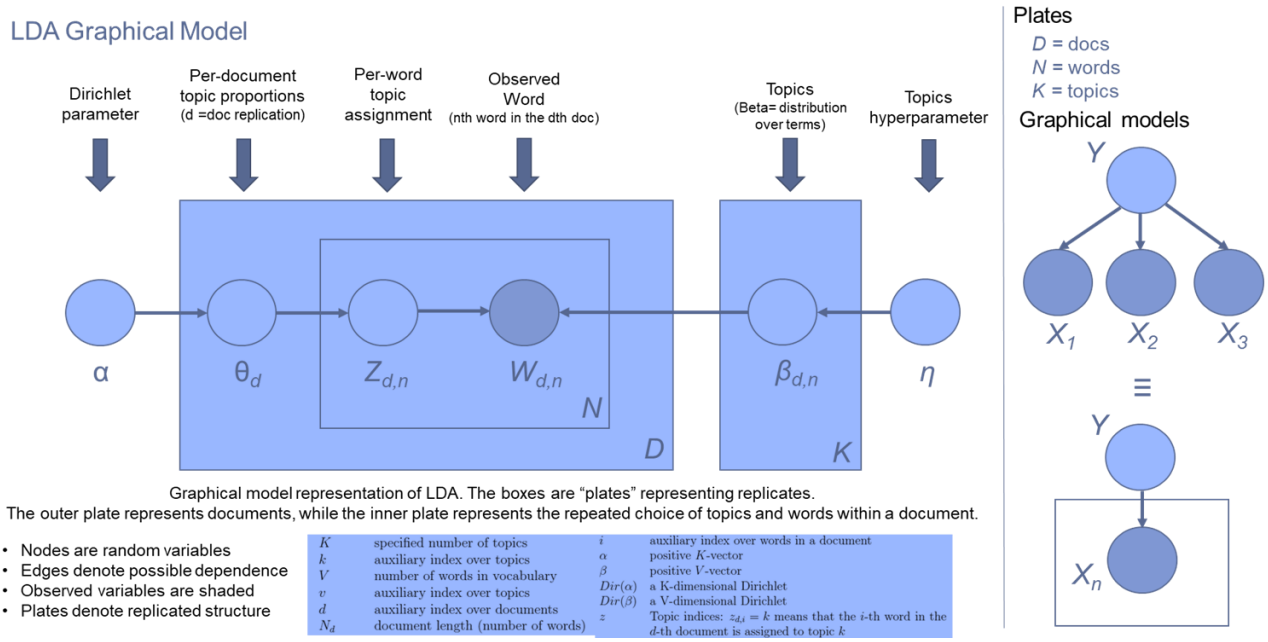


Figure 13: The Graphical Model corresponding to LDA(Source Blei. et. al. (2012)

Computing the conditional distribution (posteriors) of the topic structure given the observed documents.

$$p(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{w}, \alpha) = \frac{p(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{w}|\alpha)}{p(\boldsymbol{w}|\alpha)}$$

- $p(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{w}|\alpha)$: the joint distribution of all the random variables, that can be computed
- $p(\boldsymbol{w}|\alpha)$: the marginal probability of observations (the probability of seeing the observed corpus under any topic model) is intractable.

In theory, $p(\boldsymbol{w}|\alpha)$ is computed by summing the joint distribution over every possible combination of $\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{z}$, which is exponentially large.

An example of the LDA process is as below.

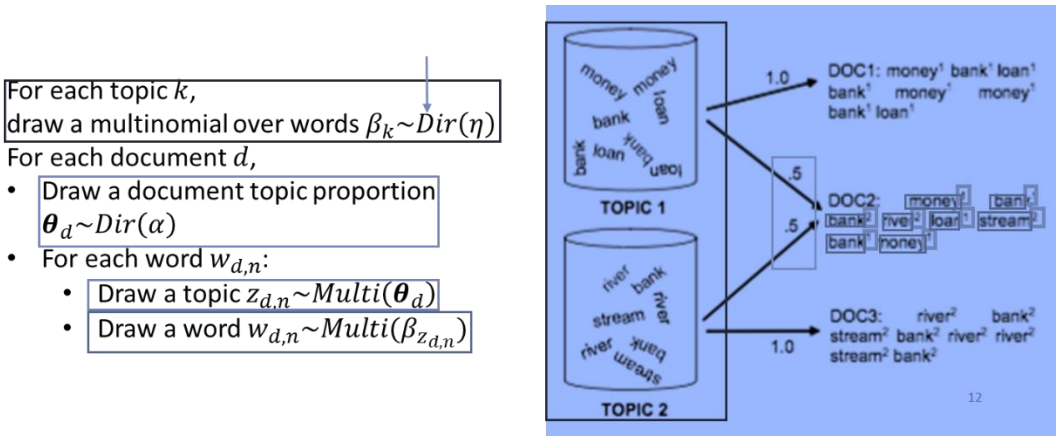


Figure 14: Example LDA

3.6.5 Approximation Methods

There are two main methods for approximation in LDA.

- ***Sampling-based algorithms*** attempt to collect samples from the posterior to approximate it with an empirical distribution.
- ***Variational algorithms*** posit a parameterized family of distributions over the hidden structure and then find the member of that family that is closest to the posterior.

3.6.5.1 In Sampling-based algorithms, Gibb's sampling

Gibbs sampling is an algorithm for sequentially sampling conditional distributions of variables, whose state distribution converges to the actual long-term distribution. We will presume to know the matrices Θ and Φ. We will slowly change these matrices and arrive at a response that maximizes the probability of the data we have. We will do this word by word by changing the theme assignment of a word. We will assume that we don't know the topic assignment of the given word, but we know the assignment of all other words, in the text, and we will try to infer what topic will be assigned to this word. We are trying to find the conditional probability distribution of a single word's topic assignment conditioned on the rest of the topic assignments. Ignoring all the mathematical calculations. We will obtain a conditional probability equation that

looks like this for a single word w in document d that belongs to topic k.

Gibb's sampling algorithm for LDA is given as below:

The hidden variables in our model are $z_{m,n}$ i.e., the topics that appear with the words of the corpus $w_{m,n}$. Gibbs sampler runs a Markov chain as per below pseudo code that uses the full conditional $p(z_i|z_{\neg i}, w)$ in order to simulate $p(z|w)$.

```
Initialization
zero all count variables, n_m^k, n_m, n_k^t, n_k
for all documents m ∈ [1, M] do
    for all words n ∈ [1, Nm] in document m do
        sample topic index z_{m,n}=k ~ Mult(1/K)
        increment document–topic count: n_m^k + 1
        increment document–topic sum: n_m + 1
        increment topic–term count: n_k^t + 1
        increment topic–term sum: n_k + 1
    end for
end for
Gibbs sampling over burn-in period and sampling period
while not finished do
    for all documents m ∈ [1, M] do
        for all words n ∈ [1, Nm] in document m do
          for the current assignment of k to a term t for word w_{m,n}:
          decrement counts and sums: n_m^k − 1; n_m − 1; n_k^t − 1; n_k − 1
           multinomial sampling acc. to Eq. 1
        decrements from previous step):
        sample topic index k ~ p(z_i|z_¬i, w)
        use the new assignment of z_{m,n} to the term t for word w_{m,n} to:

    end for
end for
check convergence and read out parameters
```

if converged and L sampling iterations since last read out then

the different parameters read outs are averaged.

read out parameter set Φ according to Eq. 2

read out parameter set Θ according to Eq. 3

end if

end while

Fig. 14. Gibbs sampling algorithm for LDA

$$p(z_i = k \mid \vec{z} - \mathrm{i}, \vec{w}) = \frac{n_{k,-\mathrm{i}}^{(t)} + \beta_l}{\sum_{t=1}^{V} n_{k,-i}^{(t)} + \beta_l} \cdot \frac{n_{m,-\mathrm{i}}^{(t)} + \alpha_k}{\left[\sum_{k=1}^{K} n_m^{(k)} + \alpha_k\right] - 1}$$

$$p(\vec{\vartheta}_m \mid \mathcal{M}, \vec{\alpha}) = \frac{1}{Z_{\vartheta_m}} \prod_{n=1}^{N_m} p(z_{m,n} \mid \vec{\vartheta}_m) p(\vec{\vartheta}_m \mid \vec{\alpha}) = \mathrm{Dir}\left(\vec{\vartheta}_m \mid \vec{n}_m + \vec{\alpha}\right),$$

$$p(\vec{\varphi}_k \mid \mathcal{M}, \vec{\beta}) = \frac{1}{Z_{\varphi_k}} \prod_{\{i_z = k\}} p(w_i \mid \vec{\varphi}_k) p(\vec{\varphi}_k \mid \vec{\beta}) = \mathrm{Dir}\left(\vec{\varphi}_k \mid \vec{n}_k + \vec{\beta}\right)$$

3.6.5.2 Variational algorithms are a deterministic alternative to sampling-based algorithms.

Variational methods provide a deterministic alternative to Markov Monte Carlo Chain (MCMC) methods. In variable approximations, a general class of distributions is defined. In this class of distributions, by utilizing optimization methods, it is determined which distribution is closer to the posterior distribution of interest. When this approximate distribution is obtained, it is then treated as the posterior distribution, and inference is made from this approximate distribution. Variable approximations are largely based on the class of distributions specified at the beginning. The more generic the class, the more complex the optimization problem. The smallest in the class, the easiest is the optimization problem. This is because variational methods are not exact methods.

For many interesting distributions, the marginal likelihood of the observations is difficult to efficiently compute. Hence, the goal here is to optimize the variational parameters to make as tight as possible and this is achieved with below three steps.

- Theorize a parametrized family of distributions over the hidden structure and then find the member of that family that is close to the posterior.
- The inference problem is transformed to an optimization problem.
- Coordinate ascent variational inference algorithm for LDA.

3.6.6 Model Evaluation:

The evaluation of the topic model is the process of assessing the ability of a topic model to do what it is intended to do. The evaluation of the topic model is an important aspect of the topic modelling process. This is because topic modeling does not provide guidance on the quality of product topics. The template is used for a more qualitative task, such as exploring semantic topics in an unstructured corpus, so assessment is more difficult. The perplexity and coherence of the subjects are used to evaluate the topic models using LDA.

3.6.6.1 Perplexity

A traditional metric for evaluating topic models is the 'held out likelihood'. This is also referred to as 'perplexity'. Perplexity is a statistical measure of how well a probability model predicts a sample. Perplexity is calculated by splitting a dataset into two parts—a training set and a test set. The idea is to train a topic model using the training set and then test the model on a test set that contains previously unseen documents (ie. Held out documents). The likelihood is usually calculated as a logarithm, so this metric is sometimes referred to as the 'held out log-likelihood'. As per models (Bao & Datta, 2014; Blei et al., 2003), Perplexity is the most typical evaluation of LDA. Perplexity measures the modeling power by calculating the inverse log-likelihood of unobserved documents (a decreasing function). Higher likelihood represents the best model.

Better models have lower perplexity, suggesting less uncertainties about the unobserved document.

$$perplexity(D_{test}) = \exp\left\{-\frac{\sum_{d=1}^{M} \log p\,(\boldsymbol{w}_d)}{\sum_{d=1}^{M} N_d}\right\}$$

A lower perplexity score indicates better generalization performance. In essence, since perplexity is equivalent to the inverse of the geometric mean, a lower perplexity implies data is more likely. As such, as the number of topics increases, the perplexity of the model should decrease.

The perplexity metric is a predictive metric that measures how well a topic model predicts a test set. In many topic model evaluations, around 80% of the corpus is a training set while the rest is the test set. This metric is natural for topic models, but it may not provide good results for human interpretation. In a recent study by Jonathan Chang and colleagues, researchers found that perplexity failed to convey the coherence of a particular topic. Statistical methods such as perplexity are useful for topic model evaluation, but aren't as reliable as human judgment. Human judgment is the best method, but it is expensive and time-consuming. Nonetheless, it is a good idea to evaluate topic models using coherence, as it is versatile and inexpensive. It's also important to know the domain in which your topic model is being evaluated to ensure its accuracy.

3.6.6.2 Model Evaluation: Topic Coherence

Topic coherence refers to the similarity in meaning of group of topics. The topic model is based on human-interpretability. This is an important step in the process of using topics in real-life conversations. The next step is to choose a metric based on the word coherence. When evaluating topic models, the evaluation process is not always as straightforward. It is possible to determine the coherence score of a document by measuring the number of successful classifications and exploration. However, this can be time-consuming and expensive. As such, we

should consider the various limitations of coherence in a conversation. It is defined as the average or median of pairwise word similarities formed by top words on a given topic. These measurements help distinguish between topics that are semantically interpretable topics and topics that are artifacts of statistical inference. Topic coherence evaluates the semantic nature of the learned topics. It measures the semantic similarity among the top keywords for a topic. Topic coherence a topic $\boldsymbol{\beta}_k$ is calculated by:

$$\text{coherence}(\boldsymbol{\beta}_k) = \sum_{(w_i, w_j) \in V_n} \text{score}(w_i, w_j)$$

(where V_n is the top n keywords of the topic $\boldsymbol{\beta}_k$)

Typical scores range between 3.0 -6.5, but values as low as 0.0 and higher than 10.0 are not uncommon.

- C_v measure is based on a sliding window, one-set segmentation of the top words and an indirect confirmation measure that uses normalized pointwise mutual information (NPMI) and the cosine similarity
- C_p is based on a sliding window, one-preceding segmentation of the top words and the confirmation measure of Fitelson's coherence
- C_uci measure is based on a sliding window and the pointwise mutual information (PMI) of all word pairs of the given top words
- C_umass is based on document cooccurrence counts, a one-preceding segmentation and a logarithmic conditional probability as confirmation measure
- C_npmi is an enhanced version of the C_uci coherence using the normalized pointwise mutual information (NPMI)
- C_a is baseed on a context window, a pairwise comparison of the top words and an indirect confirmation measure that uses normalized pointwise mutual information (NPMI) and the cosine similarity

The uMass -14 measure is a common method to measure coherence. For a human-interpretable conversation, the good lda model should have higher coherence than the bad one. The c-v and u-mass coherence measures help to determine how much coherence is present in a conversation.

3.6.6.3 Hyperparameter Tuning

Automatic model tuning, also known as hyperparameter tuning, finds the best version of a model by running many jobs that test a range of hyperparameters on the datasets. Hyperparameter tuning is choosing a set of optimal hyperparameters for a learning algorithm. A hyperparameter is a model argument whose value is set before the learning process begins. The key to machine learning algorithms is hyperparameter tuning. Model hyperparameters can be thought of as settings for a machine learning algorithm that are tuned by the data scientist before training. In LDA model m, number of topics K is one such hyper parameter. Model parameters can be thought of as what the model learns during training, such as the weights for each word in each topic. We perform a series of sensitivity tests to help determine the following model hyperparameters:

- Number of Topics (K)
- Dirichlet hyperparameter alpha: Document-Topic Density
- Dirichlet hyperparameter beta: Word-Topic Density

We run a hyperparameter tuning program to find the optimum number of K, alpha and beta using the coherence value.

3.6.7 Model visualizations

This section of the chapter explains the visualizations where we can analyze the topics created by LDA.

3.6.7.1 Dominant topic and its percentage contribution in each document:

This dominant topic for each sentence and shows the weight of the topic and the keywords. In LDA models, each document is composed of multiple topics. But typically only one of the topics is dominant. The below table shows this dominant topic for each sentence and shows the weight of the topic and the keywords.

Document_No	Dominant_Topic	Topic_Perc_Contrib	Keywords	Text
0	6	0.6788	work, contract, relate, service, legal, activity, team, role, support, development	[seek, registration, team, follow, resale, operation, concerned, party, manage, land, registrati...
1	3	0.9878	project, work, ensure, contract, review, prepare, client, contractor, management, site	[specific, include, issue, maintenance, project, contract, plan, award, management, offsite, fab...
2	7	0.9775	ensure, business, client, customer, work, develop, provide, team, support, engagement	[facility, manager, client, international, company, seek, facility, manager, candidate, backgrou...
3	9	0.9836	client, ensure, work, customer,	[person, employ, consultancy,

			sale, team, product, project, business, manage	company, develop, cut, edge, software, security, system, diverse, ...
4	2	0.5889	ensure, system, provide, service, application, legal, work, manage, customer, plan	[specialism_x, global, mobility, manager_x, career, global, mobility, service, people, organisat...
5	2	0.4689	ensure, system, provide, service, application, legal, work, manage, customer, plan	[executive, chef, full, operational, ownership, outlet, together, manager, drive, operational, a...
6	9	0.419	client, ensure, work, customer, sale, team, product, project, business, manage	[assurance, senior, associate, financial, service, senior, career, financial, service, external,...
7	2	0.9745	ensure, system, provide, service, application, legal, work, manage,	[specialism_x, manager_x, career, financial_due, diligence, practice_within, deal, transaction, ...

			customer, plan	
8	9	0.994	client, ensure, work, customer, sale, team, product, project, business, manage	[specialism_x, operation, senior, exciting, opportunity, work, government, public, sector, trans...
9	1	0.9847	sale, follow, ensure, maintain, account, market, action, marketing, hotel, client	[location, company, nestle_x, duration, month, current, university, student, marketing, position...

Table 3 : Dominant topic and its percentage contribution in each document: (Source Author)

3.6.7.2 Most representative sentence for each topic

This visualization is used when we need to get samples of sentences that most represent a given topic. This table shows the most exemplar sentence for each topic.

Topic_Num	Topic_Perc_Contrib	Keywords	Representative Text
0	0.9953	maintenance, work, ensure, equipment, perform, quality, design, contractor, project, engineer	[review, analysis, work, schedule, major, project, priority, appropriate, assignment, equipment,...

1	0.9971	sale, follow, ensure, maintain, account, market, action, marketing, hotel, client	[working_closely, colleague, achieve, personal, revenue, goal, budget, goal, property, understan...
2	0.9951	ensure, system, provide, service, application, legal, work, manage, customer, plan	[summary, description, responsible, day, day, management, logistic, function, ensure, cost, effe...
3	0.9982	project, work, ensure, contract, review, prepare, client, contractor, management, site	[job, senior, project_manager, senior, member, supervision, service, team, manage, infrastructur...
4	0.9968	customer, business, ensure, solution, client, sale, provide, team, lease, plan	[objective, lease, administrator, ensure, department, project, monitoring, phase, lease, process...
5	0.995	client, look, work, team, project, experience, contact, develop, industry, development	[specialism_x, operation, senior, globalisation, secure, affordable, energy, resource, availabil...

6	0.9939	work, contract, relate, service, legal, activity, team, role, support, development	[full, time, qualify, primary, teacher, positive, energetic, attitude, dedicate, practitioner, e...
7	0.9979	ensure, business, client, customer, work, develop, provide, team, support, engagement	[work, information, expect, travel, career, status, employment_type, regular, full, time_x, care...
8	0.9963	project, financial, review, management, prepare, work, analysis, ensure, cost, report	[motivate, team_member, deliver, unit, set, financial, management, unit, goal, link, prepare, lo...
9	0.9969	client, ensure, work, customer, sale, team, product, project, business, manage	[objective, ensure, resolve, customer, concern, sale, service, supervisor, position, require, un..

Table 4: Most representative sentence for each topic (Source Author)

3.6.7.3 Frequency Distribution of Word Counts in Documents

When working with many documents, it is required to know, how big the documents are as a whole and by topics. This report plots the document word counts distribution.

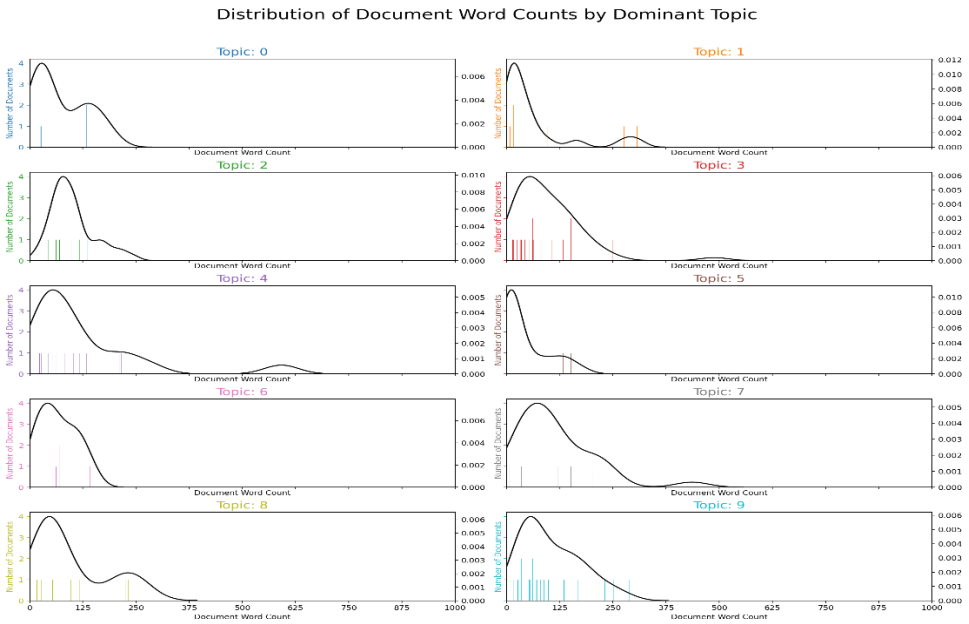


Figure 15: Frequency Distribution of Word Counts in Documents (Source- Author)

3.6.7.4 Word Clouds of Top N Keywords in Each Topic

A word cloud (also known as a tag cloud) is a visual representation of words. A word cloud with the size of the words proportional to the weight is a pleasant sight. The size of a word shows how important it is. The coloring of the topics is followed in the subsequent plots as well.

Figure16: Word Clouds of Top N Keywords in Each Topic (Source- Author)

3.6.7.5 Word Counts of Topic Keywords

When it comes to the keywords in the topics, the importance (weights) of the keywords are critical. We plot the word counts and the weights of each keyword in the same chart as below.

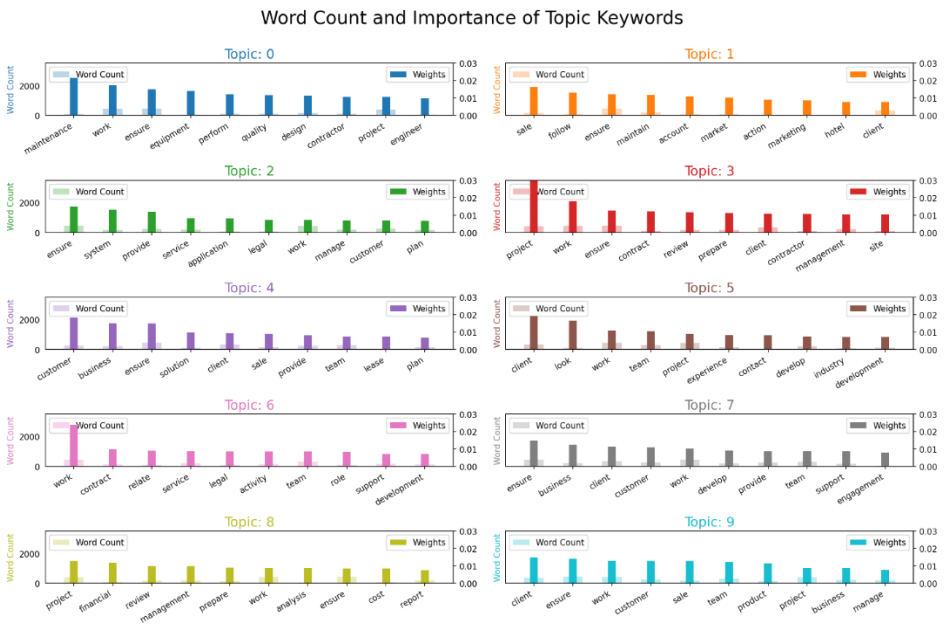


Figure 17: Word Counts of Topic Keywords (Source- Author)

3.6.7.6 Sentence Chart Colored by Topic:

In this report, each word in the document is representative of one of the four topics. Each word is attributed to the color of the enclosing rectangle and assigned to the topic of the document.

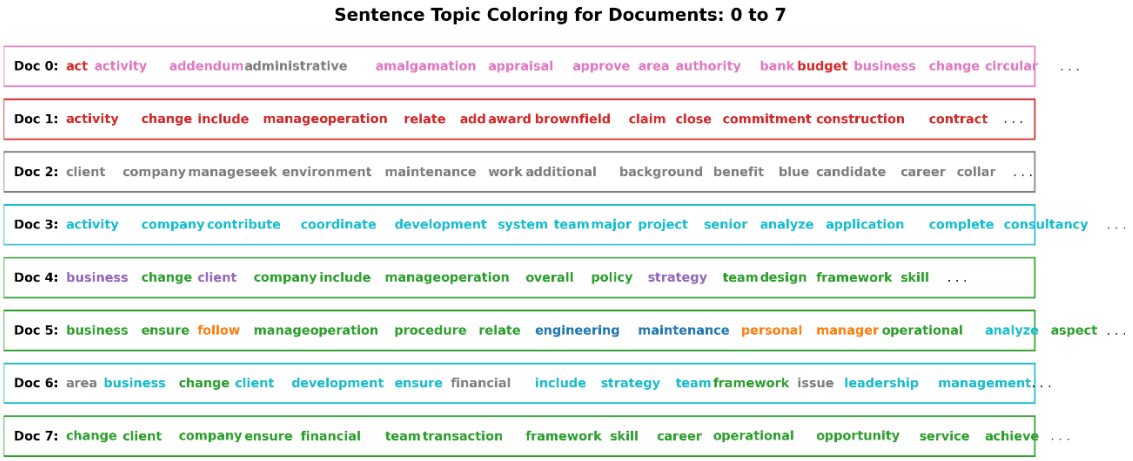


Figure 18: Sentence Chart Colored by Topic (Source- Author)

3.6.7.7 Intertopic distance map (via multidimensional scaling) Topic Model Visualization using pyLDAvis:

The intertopic distance map shows the relative frequency of words in a topic based on the number of related topics. A dark bar is used to indicate the topic-specific frequency of a word. When a word is eclipsed by a light bar, it is nearly exclusively related to the topic being examined. This data can be very useful in determining the relevance of a topic for a particular audience. Using the intertopic distance map can help marketers determine what topics are relevant for a particular audience. This diagram is a visualization of the statistical proximity of topics. Topics are represented as circles, whose size indicates their relative statistical weight. By clicking on a circle, the user can select a specific topic. The user can also search for a topic by typing the number into the search field. A further feature of the tool is the "intertopic distance map (multidimensional scaling)." This feature can give the user a clear idea of the proximity of various topics to one another. In addition to this visualization, the LDAvis package also provides interactive web-based visualization of topics. LDAvis (Sievert & Shirley, 2014) allows users to examine the terms associated with individual topics. It is a computer package that extracts information from a fitted LDA topic model and generates an interactive web-based visualization. The visualization consists of two basic parts. The left panel depicts topics as circles in a two-dimensional plane, with each circle containing its own area, representing the overall prevalence of a topic. Avis: A method for visualizing and interpreting topics

The area of these topic circles is proportional to the number of words that belongs to each topic across the dictionary. PyLDAVis is used to visualize the information contained in a topic model. It helps interpret the topics in a topic model that has been fitted to a corpus of text data. Each bubble represents a topic.

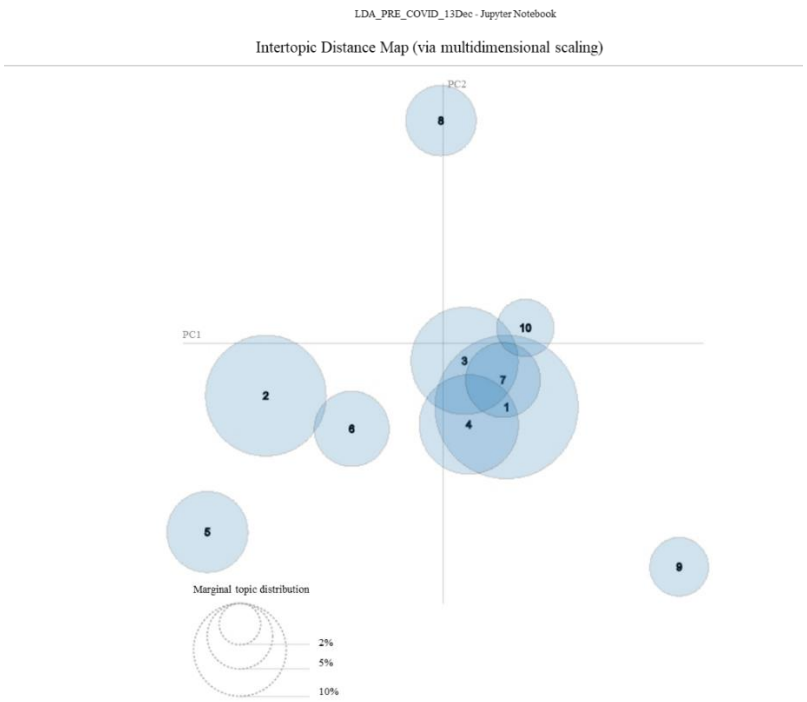


Figure 20: Intertopic distance map (Source- Author)

The larger the bubble, the higher percentage of the number of words in the corpus is about that topic. The marginal topic distribution can be interpreted as the “importance” of each topic for the whole corpus. The intertopic distance map is a visualization of the topics in a two-dimensional space. The area of these topic circles is proportional to the number of words that belongs to each topic across the dictionary. Blue bars represent the overall frequency of each word in the corpus. If no topic is selected, the blue bars of the most frequently used words will be displayed.

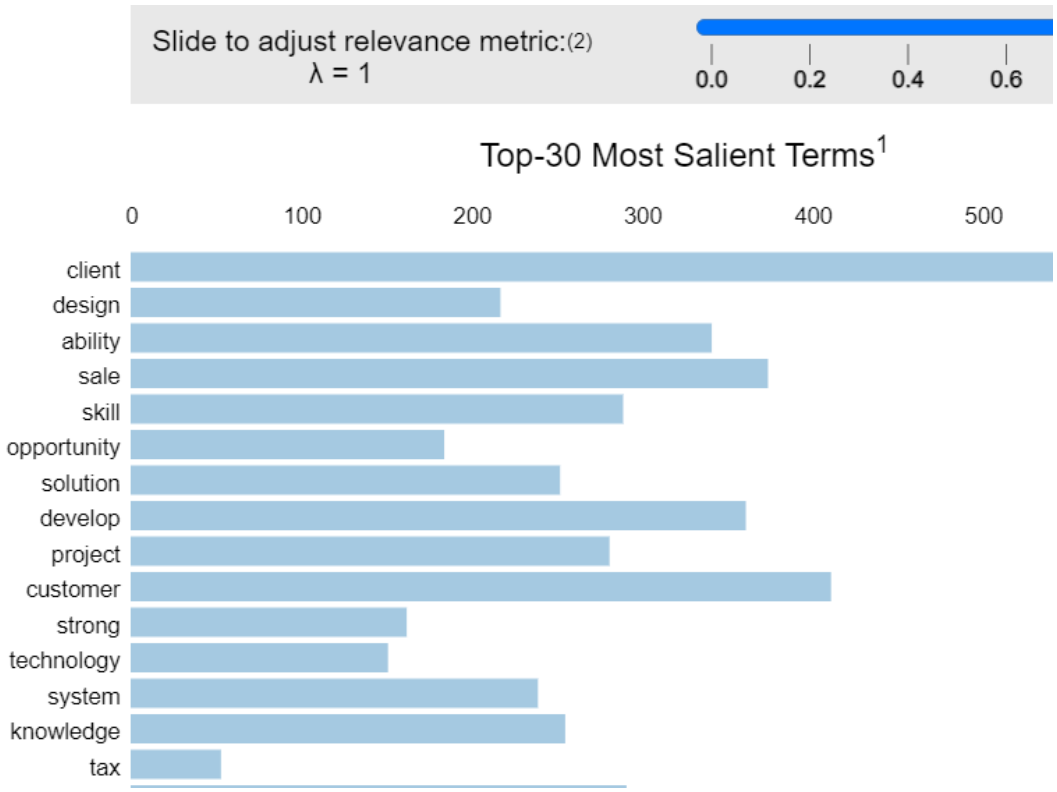
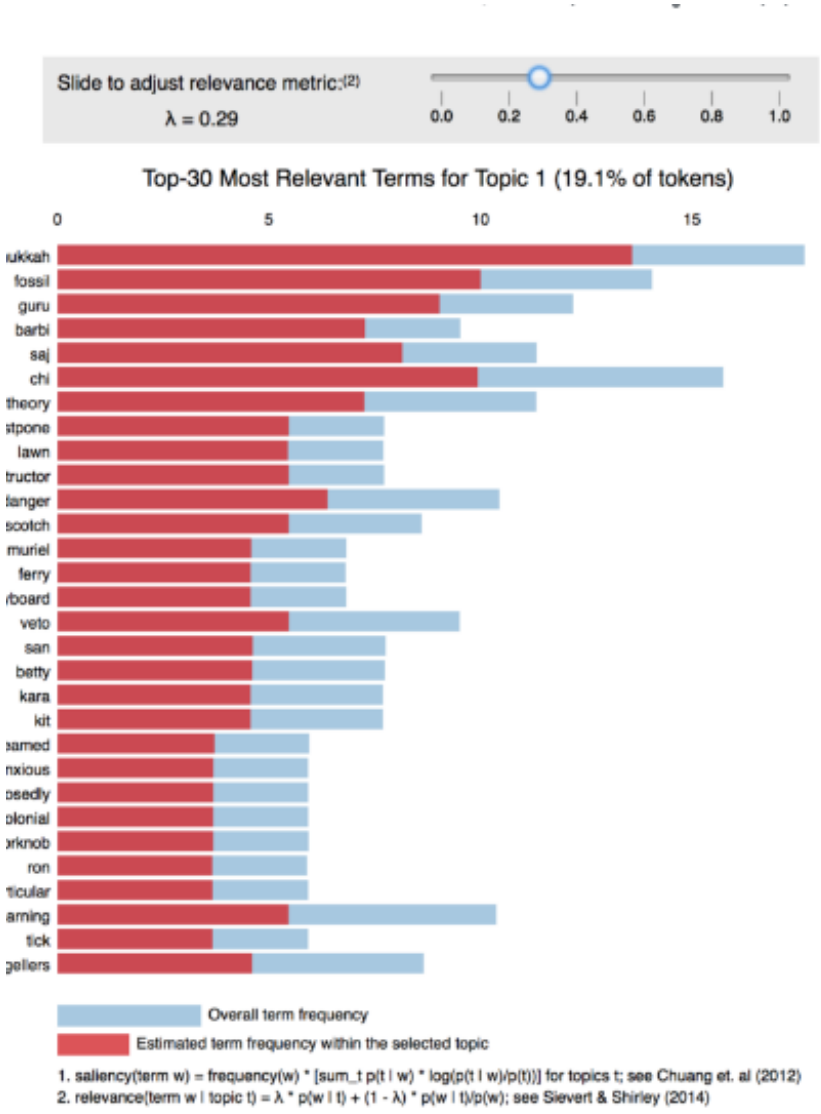


Figure 20: Overall word frequency (Source- Author)

The below visualization shows the top 30 most relevant words per topic the blue shaded bar represents the occurrence of the word in all reviews and the red bar represents the occurrence of the word within the selected topic. On top of it,we see a slide to adjust the relevance metric λ (where $0 \leq \lambda \leq 1$) and $\lambda = 1$ tunes the visualization for the words most likely to occur in each topic, and $\lambda = 0$ tunes for the words only specific for the selected topic.



LDA is a relatively simpler and popular technique for topic modelling and thanks to pyLDAvis ((Sievert & Shirley, 2014)) visualization helps to describe the functioning principle and makes topic models more interpretable and explainable.

Chapter 04: Analysis of competency demand for banking sector in the UAE- Empirical Results

The UAE banking sector is undergoing a transformation, which also means a rapid obsolescence of existing skills and an increasing demand for new skills. In addition, the misconception of the connections between skill market demand and graduate skills is presented by both sides of the labour market and universities. It is therefore important to evaluate this gap and to forecast the skills required by companies in the future. Here we attempt to get the key competencies that are demanded as per job descriptions for the banking sector in the UAE.

4.1 Data Source

The data has been extracted from published Job Descriptions from the popular job hunt sites in United Arab Emirates in the banking sector. The process looks like below. Data Sources: There are many job posting websites in the United Arab Emirates that help getting the jobs. Also, there are a few world-renowned job sites that publish jobs in this market.

Figure 1: Data Extraction process

1. Gulftalent.com: Gulf Talent is the leading job site for professionals in the Middle East. It is used by over 9 million experienced professionals from all sectors and job categories. It

serves as the key source of both local and expatriate talent to over 9,000 of the largest employers and recruitment agencies across the region.

2. Bayt: Bayt.com is another leading job search platform in the Middle East and North Africa, connecting job seekers with employers looking to hire.
3. Naukri gulf is also a website founded in 2006 that simplifies the process of hiring the right candidates across various roles, functions and experience levels across the UAE, Qatar, Saudi Arabia, Oman, Bahrain, and Kuwait.

The input data from the above includes job offer documents that will be extracted using web scraping. Web scraping is a technique to automatically access and extract large amounts of information from a website, which can save a huge amount of time and effort. We have developed it using Python as it is one of the best web scraper languages.

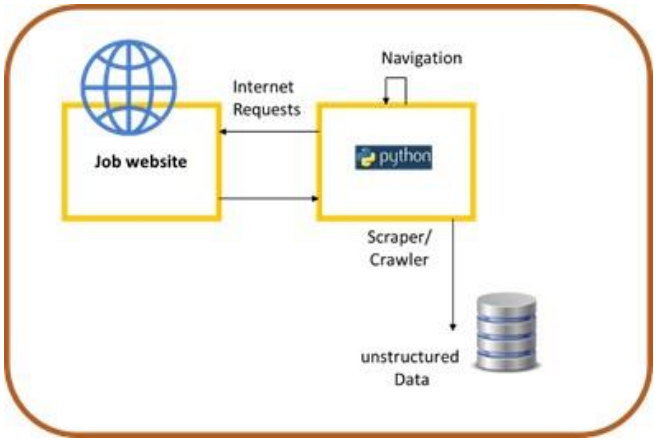


Figure 2: Web Scrapping Process (source: Author)

4.2 Topic Modelling using LDA combined data

Topic models play an essential role in exploring text data, especially with a large volume of text data, to understand the structures and groups of interest. The LDA was used for the topic modeling wherein the key topics were extracted from the bag of words. As the model follows the concept of the probability distribution of topics that describes each document, the probabilistic distribution of words can explain each topic to get a clearer vision of how the

topics are connected. This methodology has been detailed in the previous chapter.

4.2.1 Data preprocessing

Data extracted from job description sites needed to be pre-processed before they could be analyzed. This step is necessary to transform human language into a machine-readable form for further processing and analysis. Furthermore, there are some mandatory steps to clean up text from words and characters that can distort the results of experiments conducted.

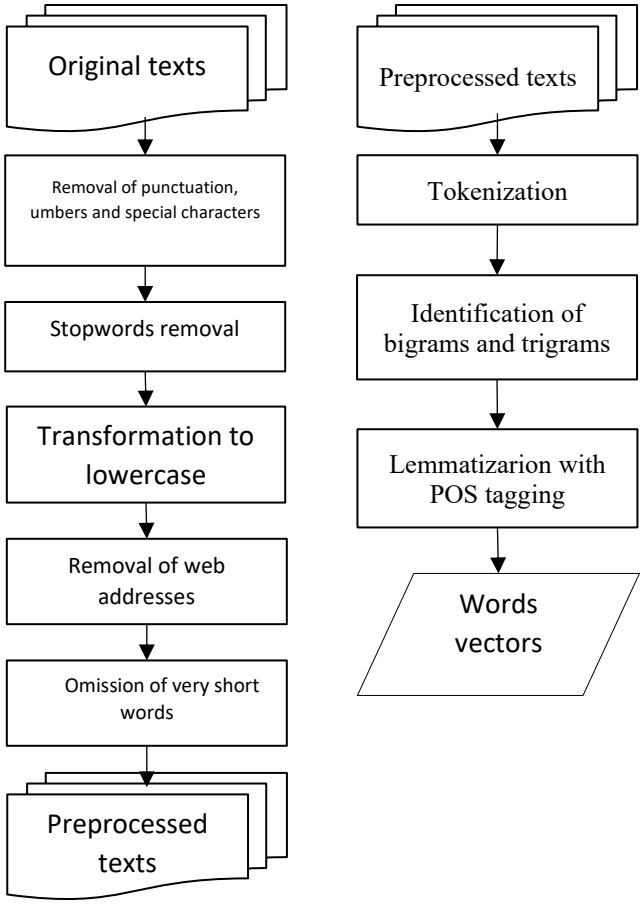


Figure 3: Data Preprocessing Process (source: Author)

- Removal of punctuation, numbers, and special characters Also, numbers and special characters without the context of surrounding text does not add any extra information.
- Stopwords removal – a lot of words in texts are common ones that have no substantive meaning. Eliminating them allows to reduce the size of analyzing corpus without loss of information. For stopwords removal list of words from nltk. Corpus library was used.
- The transformation of all characters to lowercase – this operation prevents from treating the same word but written once in lowercase and next time in upper case or in mixed form as separate entities.
- Removal of all hyperlinks if any
- Tokenization – division of each text into words. At this stage the order of words is preserved, however, due to lack of punctuation and transformation of all texts to lowercase, there is no information about syntax (i.e. Sentences).
- Identification of bigrams and trigrams – two- or three- word phrases were identified in texts using genism libraries for Python. This operation is based on dictionaries with the most common phrases constructed of two or three terms.
- Lemmatization with part of speech tagging – each token had been lemmatized using space (pretend pipeline design for text analysis). After lemmatization only nouns, verbs, adjectives and adverbials had been kept for further analysis

After preprocessing data is taken to the next step for topic modeling.

4.2.2 Topic modeling results for combined data.

We extracted and compared pre covid then post COVID as well. In this section, we analyze the combined Job descriptions. We start with hyper parameter tuning to understand the best parameters to build the models.

4.2.2.1 Hyperparameter Tuning for LDA:

We perform a series of sensitivity tests to help determine the following model hyperparameters in the model.

- Number of Topics (K)
- Dirichlet hyperparameter alpha: Document-Topic Density
- Dirichlet hyperparameter beta: Word-Topic Density

We choose the optimized parameters to build the final topic model. We run all possible numbers of topics from 4 to 10

ID	Validation_Set	Topics	Alpha	Beta	Coherence
1	100% Corpus	6	0.01	0.01	0.281022849
2	100% Corpus	6	0.31	0.01	0.281022849
3	100% Corpus	6	0.61	0.01	0.281022849
4	100% Corpus	6	0.91	0.01	0.281022849
5	100% Corpus	6	symmetric	0.01	0.281022849
296	100% Corpus	4	0.01	0.91	0.328808683
297	100% Corpus	4	0.61	0.91	0.328808683
298	100% Corpus	4	0.91	0.91	0.328808683
299	100% Corpus	4	symmetric	0.91	0.329096775
300	100% Corpus	4	asymmetric	0.91	0.329096775

Table1: Hyperparameter tuning for LDA (Source Author)

The parameter that results in maximum coherence is for number of topics = 4, alpha = asymmetric and beta =0.91.

4.2.2.2 Topic Model for Pre COVID-JDs

The LDA model above is constructed with 8 different topics according to hyper optimized settings. The model also shows the percentage of each document that deals with each topic. We obtain 4 different topics where each topic is a combination of keywords, and each keyword contributes a certain weighting to the topic.

Segments	Topic Models (Combined JDs)
Topic: 0	• project: 0.014 • management: 0.010 • report: 0.008 • support: 0.007 • review: 0.007
Topic: 1	• client: 0.016 • sale: 0.08 • customer: 0.008 • product: 0.007 • develop: 0.006
Topic: 2	• customer: 0.012 • maintain: 0.08 • report: 0.005 • account: 0.005 • provide: 0.005
Topic: 3	• accounting: 0.005 • product: 0.005 • lease: 0.004 • market: 0.003 • candidate: 0.003

Table2: Topic Clustering on JDs for the UAE banking sector pre covid (Source Author)

The key findings of term clustering of job descriptions have been performed on pre covid and this analysis results in various focus areas of hiring in the banking sector. The key categories pre covid period is the direct sales, service, consulting, PMO, Legal, IT and IT PMO etc.

4.2.2.3 Word Count and Weights of Topic Keywords

We create the word cloud with the size of the words proportional to their importance in every topic.

Figure 3: Word Count and Weights of Topic Keywords (Source Author)

4.2.2.4 Word Count and Importance of Topic Keywords:

When it comes to the keywords in the topics, the importance (weights) of the keywords matter.

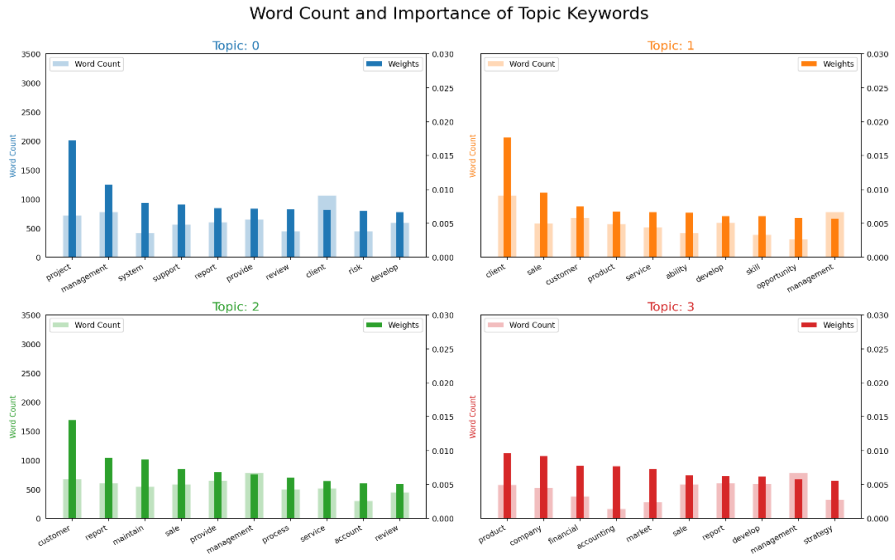


Figure4: Word Count and Importance of Topic Keywords (Source Author)

4.2.2.5 Inter topic Distance Map (via multidimensional scaling):

As per the diagram below, we identified four topics in the descending size. The further the bubbles are away from each other, the more different they are. Here all 4 bubbles are quite far hence, it shows that these four topics are very different than each other.

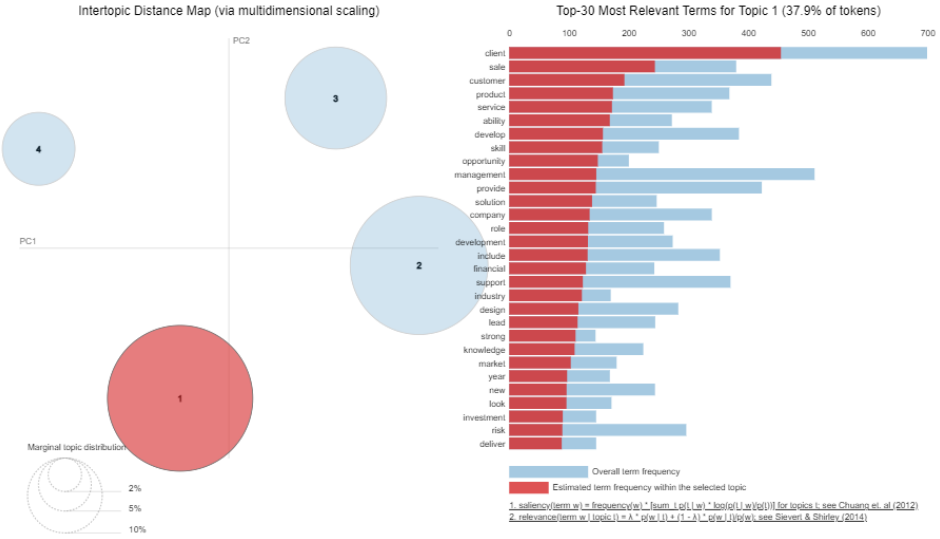
Figure 5: Inter topic Distance Map (Source Author)

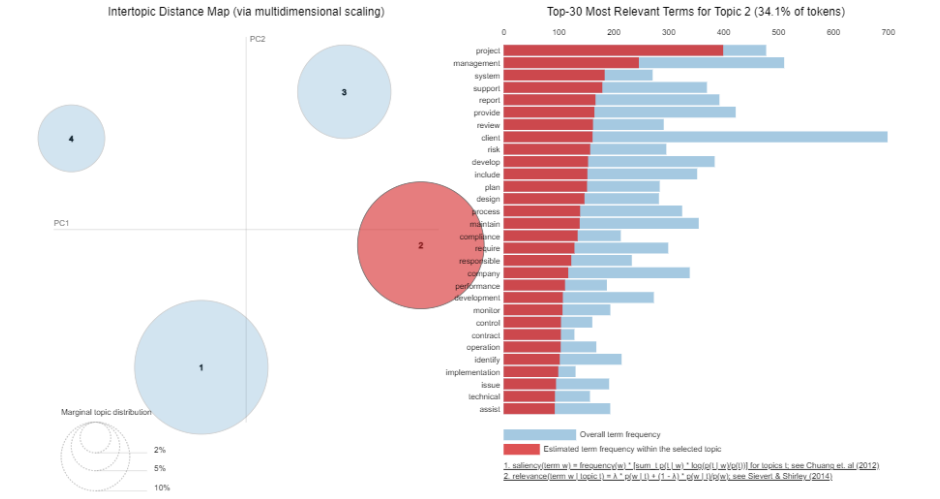
We also created intertopic distances for more topics from five to eight, as well, but this resulted in overlapping clusters.

Figure 6: Inter topic Distance Map for more topics (Source Author)

4.2.2.6 Visualization of the four topics:

Each bubble represents a topic. The larger the bubble, the higher percentage of the number of words in the corpus is about that topic. Blue bars represent the overall frequency of each word in the corpus. If no topic is selected, the blue bars of the most frequently used words will be displayed. Red bars give the estimated number of times a given term was generated by a given topic.


Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 1 (37.9% of tokens)
0
100
200
300
400
500
600
700
client
sale
customer
product
service
abillity
develop
skill
opportunity
management
provide
solution
company
role
development
include
financial
support
industry
design
lead
strong
knowledge
market
year
new
look
investment
risk
deliver
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)


Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 2 (34.1% of tokens)
0
100
200
300
400
500
600
700
project
management
system
support
report
provide
review
client
risk
develop
include
plan
design
process
maintain
compliance
require
responsible
company
performance
development
monitor
control
contract
operation
identify
implementation
issue
technical
assist
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

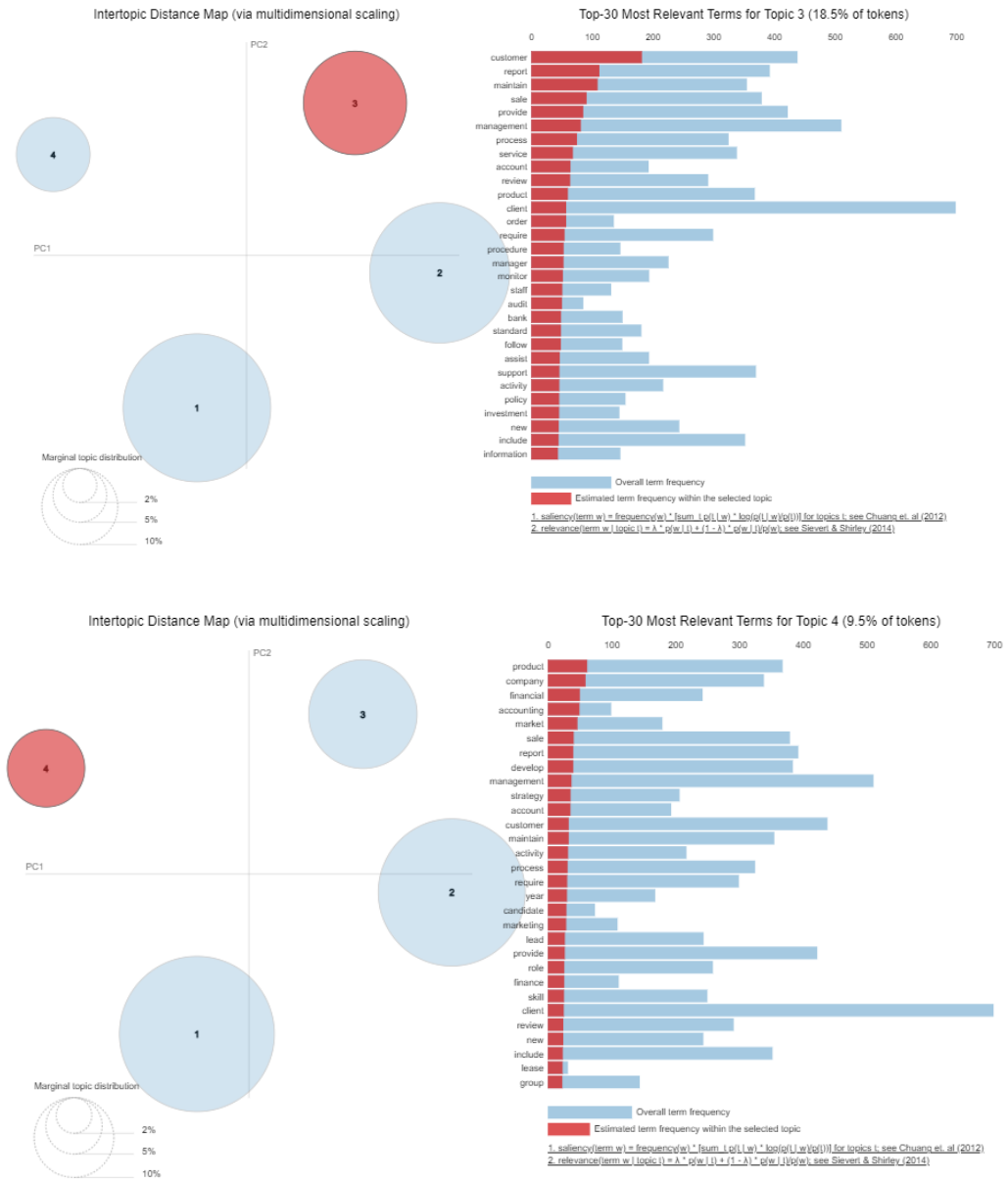


Figure 7: distribution and visualization of 4 Key topics (Source Author)

4.2.2.7 t-SNE clustering of topic models:

Tyson, (t-distributed stochastic neighbor embedding) for visualizing high-dimensional data by giving each data point a location in a two or three-dimensional map. It helps in understanding

high-dimensional data and project it into low-dimensional space. The t-SNE plot highlights clusters occurring in the original high-dimensional data. t-SNE is an effective dimensionality reduction method that is popular for topic model analysis. It is especially useful for modeling topics in text data sets. t-SNE has several key features that distinguish it from other dimensionality reduction methods. One of the most important is its tunable parameter, perplexity. t-SNE is a nonlinear dimensionality reduction method that was first introduced by van der Maatens and Hinton in 2008. It allows you to project high-dimensional data into lower-dimensional space, typically the 2D plane. Unlike other methods, t-SNE retains the structure of the original data. This makes it a powerful tool for visualising complex datasets, especially those with non-linear relationships between observed features and target classes. For example, t-SNE can often uncover patterns in scRNA-seq data that other techniques cannot. It can also detect clusters in images with low resolution or poor color coding, or reveal the relationship between topics in a set of text documents. However, t-SNE is not without its limitations and can sometimes produce misleading results. It is important to understand some of these limitations, as well as how to interpret t-SNE plots correctly.

T-SNE is not deterministic in the sense that every run of the algorithm will produce a different output. This is because it uses a stochastic probability distribution to model the relationships between points in high-dimensional space. The probability of a point being close to another is modeled as the probability of its neighbors (the set of points closest to a given point). This probability is then mapped into a low-dimensional space, called the embedding. This procedure tries to maintain the neighborhood of the original point in the lower-dimensional space. To do this, a probabilistic mapping of the Gaussian distribution to a Student's t-distribution is used. The t-distribution has fatter tails than the Gaussian distribution, which helps spread points more evenly in the embedding space. However, this is not always possible, especially when the data has a high degree of clustering. In this case, the perplexity parameter, which sets the number of points to guess at each neighbor, has to be fine-tuned to balance global and local structures.

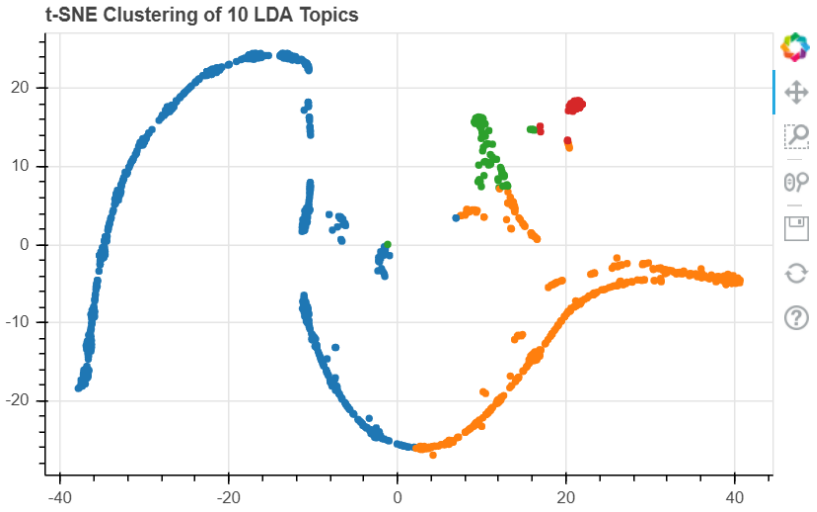


Figure 7: t-SNE clustering of topic models (Source Author)

it works by generating a probability distribution of Euclidean distances between points in the input data space, called perplexity. Its value is controlled by a user. Low perplexity values force t-SNE to focus on the local structure of the data while larger ones take global structure into account. This is what makes t-SNE especially useful when dealing with complex manifold structures.

4.2.2.8 Sentence Topic Coloring for Documents:

This visualization helps as each word is colored in the given documents by the topic id it is attributed to.

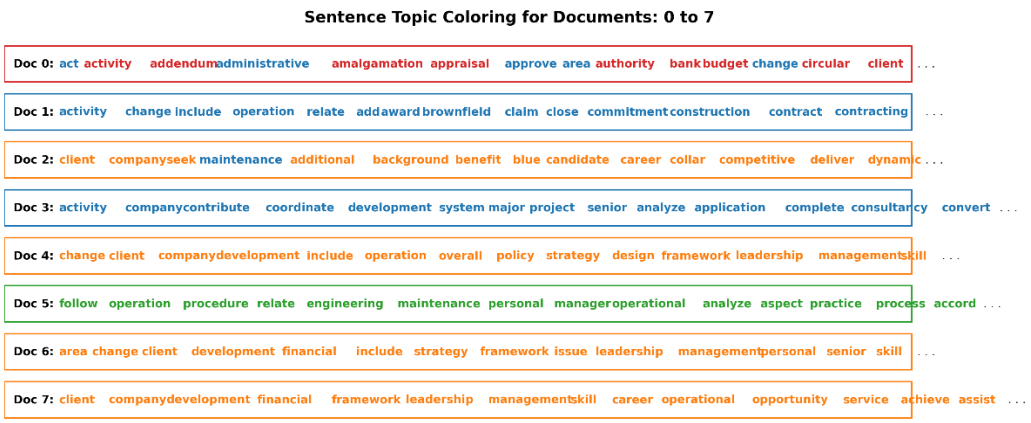


Figure 8: the most exemplar sentence for each topic (Source Author)

4.2.2.9 Distribution of Document Word Counts by Dominant Topic:

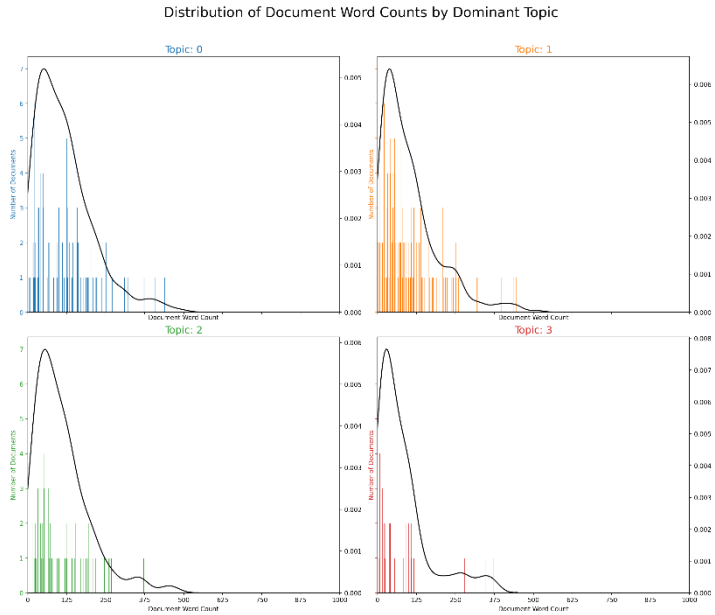


Figure 12: The most exemplar sentence for each topic (Source Author)

4.2.2.10 Topic Distribution by Dominant Topics c:

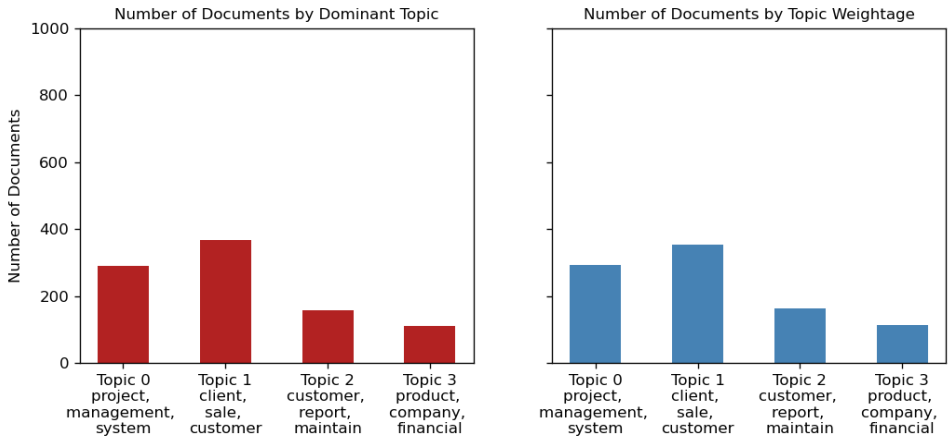


Figure 12: the most exemplar sentence for each topic (Source Author)

4.2.2.11 Distribution of Document Word Counts:

To know how big the documents are as a whole and by topic, we plot the document word counts distribution.

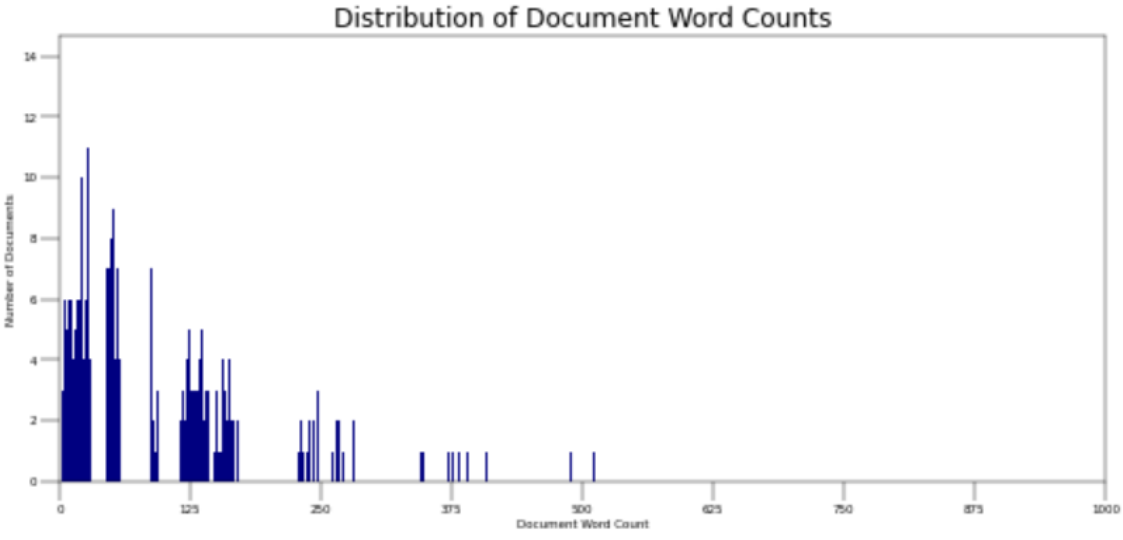


Figure 4: Distribution of Document Word counts (Source Author)

4.2.3 Comparing Topic Distribution pre and post covid:

We have compared pre and post covid dominant topics, and there is a here are the noticeable differences among key dominant topics.

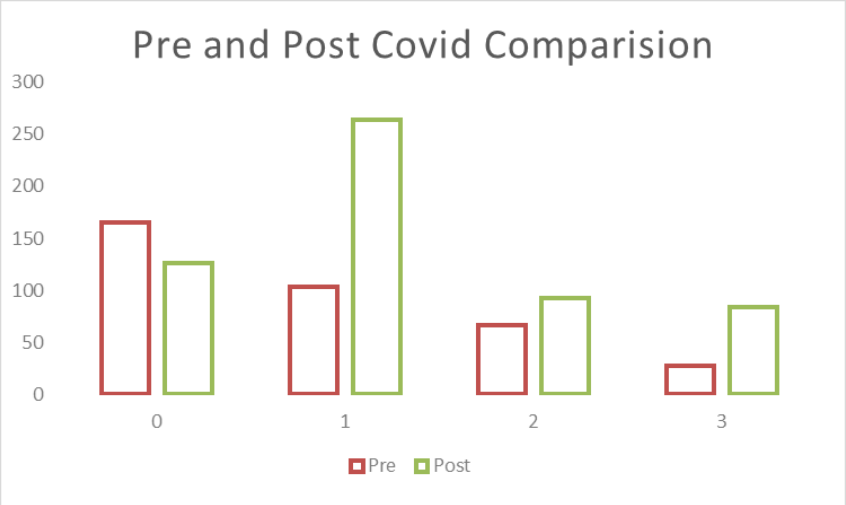


Figure 4: pre and post covid topic comparison (Source Author)

This analysis results in various focus areas of hiring in the banking sector. The key findings after term clustering of job descriptions have been performed on pre covid and post covid data. One of the topics that got reduced in post covid job descriptions were focusing on project management, reporting, support staff etc. Post covid these topics got reduced significantly. Back-office has been made bare thin and outsourced and investments reduced. One area that has reversed the trend from another topic category focuses on clients and sales. This made back-office and middle-office staff not being required much during fill-back period. Risk management has been a growing category where there has been a focus in hiring in banking sector in the UAE.

There have been several interesting works in the same direction. Lee, Sangheon et. Al. (2021) estimated expected that the income of informal workers will decline in the first month of the crisis by 60 per cent globally. In Africa and the Americas, the drop is even over

80 per cent. Bell, David NF (2020)'s survey estimated that a third of workers in Canada and the US report that they have lost at least half of their income due to the Covid-19 crisis. Barker, Nathan et al. (2020) created pre- and post-COVID panel datasets for three populations in Bangladesh and Nepal, leveraging experimental and observational variation in prior migration dependence and report 25 per cent greater declines in earnings. Alexander W. Bartik et.al (2020) found that impact was driven by low-wage services. Bikse, Veronika, et al.(2021) found that The work and learning of many individuals have moved to the digital environment. To use digital technologies, employees need to acquire new knowledge and skills.

4.2.4 Conclusion and Discussions:

Competencies demanded by companies are changing faster now than before (Mckinsey 2020). AI systems mainly NLP, can be used to identify the emerging trends in the hiring by industry. Governments, University and private institutes have to play an active role in creating awareness of emerging technology trends. This unique application of topic modeling can be used for further representation of crucial patterns, changes in the labor market in the banking sector.

Chapter 5: Proposal of the network model of selected aspects of the labor market

5.1 Use of graph models in labor markets: network models for labor market

There have been various applications of the graph models in the labor market. Let us start with a few examples. Using graph-based methods for competency analysis can help us to identify which employees are likely to succeed in each role. This method has been used to assess employee performance and to make more informed decisions about training sessions and promotions. Other applications of Graph data models where it has been used to rank people by their relationships with other employees. By using this method, we can easily determine the availability of competencies for different roles and teams within an organization. Graph data models are useful for analyzing internal organizational structure. Nodes represent employees and roles. Data relationships between nodes show the length of tenure of positions, activities performed, and employee evaluations. A job recommendation model can be modeled by using a graph data model and elements of the algebra competency management model. Wang C. (2021) analyzed the performance of different employees and build a model to help make better hiring decisions they. Extracted the latent interpretable representations of the employees' competencies from their skill profiles with auto-encoding variational inference-based topic modeling. Then developed an effective demand recognition mechanism for learning the personal demands of career development for employees.

The labor perspective in terms of demand of competencies is presented in different research conducted in the relevance to various countries and fields of economy. In this regard it is worth mentioning the studies devoted to analysis of matching labor market demands with competencies of graduates. One of the methodological approaches is aimed to analyze the list of competencies presented in job advertisements (Suarta I. M. et al, 2018; Olawale, 2015; Dunbar K., et al., 2016; Messum D. et al., 2016). Another method is to recognize competence requested by companies by ontology-guided job market demand analysis (Sibarani, E. et al., 2017). Lula, P. et al., (2017) performed the analysis of job offers published online in the portal

https://www.pracuj.pl/ and build the competency co-occurrence graph for Polish labor market. He analyzed five thousand job offers from which required competencies were identified.

5.1.1 Social Networks and Job Searches:

Mark Granovetter (1973) interviewed professional, technical and management (PTM) employees from the Boston area to explore how they got their current job. Knoke (2012) shows a self-centered network where members can randomly pass redundant information to other members that will greatly benefit from this new intelligence on engineering openings. Wegener (1991) explained that people who change jobs and have low initial prestige and strong intimacy and provided the structural lever for career progression. The Granovetter study (1973) found that professional prestige is inferior to that of TMP workers, strong ties matter in their job search. Bean (1997) observed how this network helped unemployed people in China and Lin, Fu & Hsung (2001) analysed this in Taiwan.

5.1.2 Networks and Filling Job Vacancies

Harry Holzer (1987) performed network analytics application in filling job vacancies and reported that 36% of firms in his study used the method. Kalleberg et al. (1996) and Marsden & Campbell (1990) reported that more than 51% of jobs are filled through referral. Cingano et al. (2012) found that employed individuals will likely have a positive impact on the probability of their unemployed friends finding a job. Calvo-Armengol and Jackson (2004) experienced that asymmetric effects can potentially result in large discrepancies in long-term labor market outcomes arising from small temporary employment shocks.

5.1.3 Social Networks and Job Placement in multiple Countries

Wegener (1991) found that weak ties are the most helpful in landing them an ideal job in Germany. Kramarz & Skans (2014) performed network analytics for filling job vacancies and found that, young workers use their parents' strong social ties to land their first job in Sweden. The benefits of using strong ties are especially large when the unemployment rate is high or when the youth have low education or bad grades. Bin (1994) found that Chinese government controlled the laborers and assigned jobs. Lin (2001) established that through a two-step hop, a

job seeker eventually relates to the agent who may be persuaded to favor the job seeker. Such strategic mobilization of one's network contacts to facilitate one's social action is one key process in actualizing one's social capital in the Chinese labor market.

5.1.4 Role of Social Networks in Labor Market Outcomes

Filipa Reis (2015) used a new and unique dataset combining social network data from call detail records with employment information on mobile phone subscribers to study the role of information networks on job market outcomes.

5.2 The concept of the model of employers' expectations towards candidates for employment in banking sector in the UAE

To understand the employer's expectation, it is critical that we understand the job description in detail. This is because the job description is the blueprint for the employer's expectations and will help us understand how to meet them. In addition to understanding what a prospective employer wants, it is also essential that we understand the role the employer wants to hire. The job description includes the various requirements for the position. A well-written job description includes a list of skills, knowledge, and behavior needed for a certain job. Using job descriptions to understand competencies is one way to understand new skills that are in demand. Job descriptions give employees a sense of purpose, while also outlining how their role contributes to the organization's goals. We aim to build a hierarchical model to analyze the competency in demand in the banking sector in UAE.

5.2.1 General description of the proposed model:

A job description provides essential details necessary for the performance of the position. It also enables organizations to attract and recruit qualified applicants. An effective job description describes the specific duties and responsibilities of a position, as well as the reporting structure. A well-written job description includes the purpose of the role, as well as its contributions to the organization's mission. The job description contains at least 4 or 5 essential duties that are most relevant to the position. In addition, a good job description also includes the work-site layout, equipment, and the location of essential functions. In addition, the description indicates whether the job is supervisory. If so, it should be stated as a mandatory function. If not, it should be clearly

indicated that it has no supervisory function. There are various components of job descriptions (JDs). The schema of the JD captures many elements such as posting for a location, years of experience needed, competency needed etc. The typical structure of the requirement is captured in the following structure.

5.2.1.1 Job position:

A job position is a function served in an organization. It includes the daily tasks and projects to complete. Every employee has a job position that includes specific duties and responsibilities that help the organization reach its goals. JD contains the details of the job positions that can be used to analyze the demand of different functions or roles. There are many different types of banking jobs published in the UAE. From entry level to executive level, there are various JDs. Many large banks employ a range of staff members, including human resources, payroll, information technology, and secretarial and administrative support. There are also rules to work in computer technology or programming to develop applications. There are several different banking career paths to choose from. Some examples of jobs in the banking industry include investment banker, mortgage processor, loan officer, and risk manager. The financial sector is an industry that is constantly changing, and the banking industry is no exception. The financial crisis has caused a great deal of change in the banking industry. The economy is experiencing an unprecedented period of growth, and banks must continue to adapt to these changes. A bright outlook and the ability to change with it are valuable traits for a banking job. Hence, we need to dynamically explore the positions and roles published by finical services companies.

5.2.1.2 Sector

Job sectors are professional categories that, in total, describe most careers. Various examples of Job sectors are Health Care and Social Services, Leisure and Hospitality, Local Government, Retail Trade, Manufacturing, Professional and Technical Services, Administrative Services, Financial Activities.

5.2.1.3 Company

A job description should include important organization or company details that have the requirement. This helps in identifying if the demand has been from a domestic company or a MNC. As covered in chapter 1, UAE has many domestic, regional, and international banks. The JDs extracted here cover all such entities in the UAE.

5.2.1.4 Location

For our study, we categorize three locations within the UAE. The country is a federation of seven constituent monarchies: the Emirates of Abu Dhabi, Ajman, Dubai, Fujairah, Ras al-Khaimah, Sharjah, and Umm al-Quwain. Most of the requirements are in Dubai and Abu Dhabi. So, we have created 3 location segments for the locational analysis.

Locational Region
Dubai
Abu Dhabi
Others

Table 1: City of Job Postings (Source: Author)

5.2.1.5 Expectations

It defines the role and primary functions, states the duties of the position, clarifies reporting relationships and responsibilities, and outlines the key skills and abilities required.

5.2.1.6 Experience

Work experience is the experience an employee gains while working in a job, particular field or profession. JDs mention the experience in the related job or role.

5.2.1.7 Required competencies

Competencies in a job description identify the desired and required skills and behaviors needed to perform a job successfully. Competencies can identify required soft skills – for example, “attention to detail” or “fostering communication.

5.2.2 The proposal of the hierarchical model of competencies in the banking system

The hierarchy-based structure is very suitable for competency assessment for banking and other industries. We first focus on high-level competencies that are relevant to the banking industry. As the next step, we go into detailed competency framework within each high-level competency. A hierarchical model of competencies is a logical structure based on the concept of skill, with some categories being more important than others. It is important to distinguish between different types of skills in different situations, which is why the skills listed in one category may not be appropriate in a different context. Competencies are typically defined in terms of their degree of difficulty, but they may also vary from person to person. Here we present a hierarchical model of competencies that incorporates existing types of competencies that are demanded by the organizations in the banking sector.

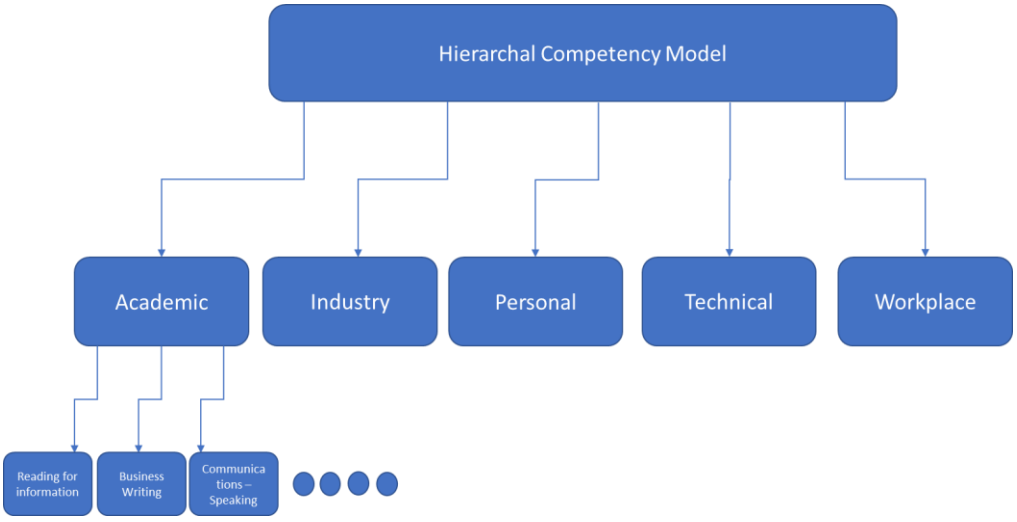


Figure 1: Hierarchical competency model (Source: Author)

This automated software tool for performing a comprehensive analysis of supply- and demand-side of a labor market in UAE for banking sector that could easily be extended to other industry sectors. This hierarchal model has the following scalability advantages:

- Hierarchical is a suitable model for identifying industry specific competencies

- It is helpful for automatic analysis (computational)
- It is Universal (it will be used in UAE, but can be used in other regions)
- It is holistic (not only competencies, but other objects related to the labor market, such as employees, employers, companies, locations
- We have created a keywords library to match every competency for mapping the JDs and competency. The keywords are extensive and based on industry practice and published Job Descriptions and competency, key words from the leading 3 websites (or jobsites).

5.2.3 Model implementation

As part of this research, we aim to implement the proposed competency model for the banking sector in the UAE. To start the process, we need to extract the data related to job descriptions from jobsites.

5.2.3.1 Ontology-based identification of fragments of offers related to the model's components

An ontology is a set of concepts that describes the relationships among concepts. It is useful to describe atomic and combined concepts, as well as final facts.

5.2.3.2 High-level competencies proposed in the banking system

Based on research and experience, the five categories that have been identified are as follows:

1. Academic competencies
2. Industry-wide competencies
3. Personal effectiveness competencies
4. Technical competencies
5. Workplace competencies

5.2.3.2.1 Academic competencies:

Academic competencies are essential success, enabling skills that improve the learner's utilization of reading, writing, mathematics, science and technology, computers, effective communication, critical thinking, and fuel career success in any industry including the banking

sector. A few attempts to circumscribe the learner characteristic domain are reflected in the works of Diperna (2000) and Elliott (2002). In their Model of Academic Competence (MAC), they defined academic competence as "a multidimensional construct consisting of the skills, attitudes, and behaviors of learners that contribute to success in the classrooms". Academic competence includes the domains of academic skills and academic enablers. According to DiPerna and Elliott (2002), "academic skills are the basic and complex skills that are the primary focus of academic instruction in elementary and secondary schools. In contrast, academic enablers are attitudes and behaviors that allow a learner to participate in, and ultimately benefit from, academic instruction in the classroom". Key academic competencies that are considered here are:

High Level Competency	Detailed Competencies
Academic competencies	Reading for information
Academic competencies	Business Writing
Academic competencies	market economics
Academic competencies	Communication listening
Academic competencies	Communication speaking
Academic competencies	Critical thinking
Academic competencies	Analytical thinking

Table 2: Academic competencies (Source: Author)

5.2.3.2.2 Industry-wide competencies

The general concept of industry competency focusses on the ability to perform tasks to the standard of performance expected in the workplace. Operators and contractors within the industry have established requirements for safety training and competence.

High Level Competency	Detailed Competencies
Industry-wide competencies	Anti-Money Laundering
Industry-wide competencies	Audit
Industry-wide competencies	Branch administration
Industry-wide competencies	Business Analyst
Industry-wide competencies	Business operations
Industry-wide competencies	Business strategy
Industry-wide competencies	Client advisory

Industry-wide competencies	Content writer
Industry-wide competencies	Customer service
Industry-wide competencies	Data Science
Industry-wide competencies	Financial reporting
Industry-wide competencies	Fintech
Industry-wide competencies	Fraud preventions
Industry-wide competencies	Human Resource
Industry-wide competencies	Information Technology
Industry-wide competencies	Insurance
Industry-wide competencies	Legal
Industry-wide competencies	legal and compliance
Industry-wide competencies	Management information systems
Industry-wide competencies	marketing
Industry-wide competencies	Organizational development and human resources management
Industry-wide competencies	Privatization, restructuring, acquisition and merger
Industry-wide competencies	products
Industry-wide competencies	Real Estate
Industry-wide competencies	Regulations
Industry-wide competencies	Risk Management
Industry-wide competencies	Sales
Industry-wide competencies	Securities commissions
Industry-wide competencies	Software
Industry-wide competencies	Wealth Management

Table 3: Industry wide competencies (Source: Author)

5.2.3.2.3 Personal effectiveness competencies:

Social and personal competencies are a set of skills to include self-awareness, self-management, social awareness, relationship skills, and responsible decision-making. These are the soft skills that are needed for students to succeed in postsecondary and career. An essential part of this competence is intentional. Without intention, there is no competence. Personal competencies, therefore, forecast certain behavior. Personal competencies enable us to understand why some people perform better at work than others.

High Level Competency	Detailed Competencies
Personal effectiveness	Interpersonal Skills
Personal effectiveness	Ethics

Personal effectiveness	Integrity
Personal effectiveness	Credibility
Personal effectiveness	Self-Management
Personal effectiveness	Reliability
Personal effectiveness	Dependability
Personal effectiveness	Time Management

Table 4: Academic competencies (Source: Author)

5.2.3.2.4 Technical competencies:

Technical competencies are behaviors directly related to the nature of training and the technical proficiency required to exercise effective control. Competency on a task requires a match between the operator's competencies and the competencies required to perform that task safely and effectively. Technical skills are the abilities and knowledge needed to perform specific tasks.

High Level Competency	Detailed Competencies
Technical competencies	Big Data Analysis
Technical competencies	Programming
Technical competencies	UI Design
Technical competencies	Social Media Management
Technical competencies	Microsoft Office
Technical competencies	Cloud
Technical competencies	SaaS

Table5: Technical competencies (Source: Author)

5.2.3.2.5 Workplace competencies

A workplace competency is a description of a required skill, attribute or behavior for a specific job used to define and measure an individual's effectiveness. They reflect the knowledge, skills, and abilities that are most relevant in today's workforce.

High Level Competency	Detailed Competencies
Workplace Competencies	Adaptability
Workplace Competencies	Flexibility
Workplace Competencies	Health and safety
Workplace Competencies	Leadership Skills
Workplace Competencies	Planning

Workplace Competencies	Problem solving
Workplace Competencies	Procurement manager
Workplace Competencies	Project management
Workplace Competencies	Teamwork
Workplace Competencies	Transportation manager
Workplace Competencies	Housekeeping manager

Table 6: workforce competencies(Source: Author)

5.2.3.2.6 Competencies building from key words

As there could be many keywords to capture the same competencies. We have added as many as possible keywords that could capture the same competencies. Here are a few examples.

High Level Competency	Detailed Competencies	Keywords
Workplace competencies	Procurement manager	Procurement activities
Workplace competencies	Procurement manager	Procurement manager
Workplace competencies	Procurement manager	Procurement specialist
Workplace competencies	Procurement manager	Sourcing new suppliers
Workplace competencies	Procurement manager	Supplier Management Officer
Workplace competencies	Project management	Facilitator
Workplace competencies	Project management	Innovation manager
Workplace competencies	Project management	PMO office
Workplace competencies	Project management	Program manager
Workplace competencies	Project management	Program Manager
Workplace competencies	Project management	Project management
Workplace competencies	Project management	Project status reports
Workplace competencies	Leadership Skills	Develop and implement strategy
Workplace competencies	Leadership Skills	Leadership
Workplace competencies	Leadership Skills	Ownership
Workplace competencies	Leadership Skills	People management
Workplace competencies	Leadership Skills	Responsibility
Workplace competencies	Leadership Skills	Accountability
Workplace competencies	Leadership Skills	Leadership Skills
Workplace competencies	Transportation manager	Logistics
Workplace competencies	Transportation manager	Routine maintenance of transportation vehicles
Workplace competencies	Transportation manager	Transportation manager
Workplace competencies	Warehouse manager	Warehouse manager

Workplace competencies	Warehouse manager	Warehousing

Table 7: Industry wide competencies (Source: Author)

5.2.3.3 Input Data Structure:

As a next step we build a matrix where rows correspond to the offers and columns correspond to competencies, position, and location. Here is a proposed structure for the same as an example.

	Job position	Sector	Company	Location	Years of experience	Competency 1	Competency 2	…	Competency M	
Job 0ffer 1										
Job offer 2										
….										
Jon offer N										

Table 8: Matrix Structure Industry wide competencies (Source: Author)

The columns of the data structure have Job position, Sector, Company, Location, Years of experience, Competency – 1 (a competency represented by a column mentioned in a job offer represented by a row) or 0 (not mentioned).

This matrix was used to build the graph for competency analysis.

5.3 Analysis of competences and relationships among them

The development of an analysis of job competencies is a powerful tool for understanding organizational needs. Many companies use job competencies to assess their workforce's capabilities as well. In addition to identifying the skills a candidate needs for a specific position, an analysis of job competencies can also identify the skills they don't possess. Using the document-competency matrix, employees can evaluate themselves against the standards set out by the organization. In addition, employers can use the analysis to help identify which candidates are best suited for the job.

5.3.1 Building a matrix of competency co-occurrence

The traditional matrix structure of competencies could be challenging to manage. Here, we recommend a detailed competency co-occurrence matrix. Co-occurrence matrix computes often occurring pair of competencies within a job description. We expand the competency co-occurrence structure with geography and other demographic information so that we can quickly and effectively understand the competence demand across markets. The competency matrix also helps us understand the company needs for the projects or job functions to be executed successfully.

Let's assume that $\boldsymbol{C}$ is a set of competencies:

$$\boldsymbol{C} = \{C_1, C_2, \dots, C_M\}.$$

In this context, a competency schema can be defined as a weighted graph represented by an adjacency matrix $\boldsymbol{G}$:

$$\boldsymbol{G} = \left[g_{ij}\right]$$
$$i, j = 1, \dots, M$$

where diagonal elements g_{ii} inform about the importance of the i-th competency, and off-diagonal elements g_{ij} describe a strength of relations between competencies C_i and C_j

The process of building of the competency schema is presented below.

We will consider a *set of competencies*:

$$\boldsymbol{C} = \{C_1, C_2, \dots, C_M\}$$

and a *set of objects* to which competencies are assigned (e.g., job offers, job positions, members of staff, candidates for employment, responders):

$$\boldsymbol{O} = \{O_1, O_2, \dots, O_N\}$$

The analysis of offers enables to define the *object-competency matrix* $\boldsymbol{M}$:

$$\boldsymbol{M} = [m_{ij}]$$
$$i = 1, \dots, N; j = 1, \dots, M$$

where $m_{ij} = 1$ if the j-th competency is mentioned in the context of the i-th object, and $m_{ij} = 0$ if the information about the j-the competency does not appear in the context of the i-the object.

We assume that two competencies C_i and C_j are connected if they appear in the context of the same object O_k. Connections between competencies are represented by the *competency co-occurrence graph* which has weighted and undirected character and is represented by the adjacency matrix $\boldsymbol{R}$:

$$\boldsymbol{G} = [r_{ij}]$$

where r_{ii} element informs how many the C_i competency is mentioned in the context of the given data set $\boldsymbol{O}$, and r_{ij} (where $i \neq j$) shows how many times competencies C_i and C_j appeared together in the context of the same object. The competency schema can be defined as a community structure G identified in the competency co-occurrence graph R.

5.3.2 Evaluation of competency importance- Number of occurrences

There are various statistics that can help in the evaluation of the relative importance of competencies. We can use the matrix of competency co-occurrence to highlight the different skills an organization needs. This matrix is helpful for both the organization and the individuals to function more efficiently, as it communicates expectations and gaps in skill-sets. Building a

matrix of competency co-occurrence helps to identify weaknesses and strengths in teams and allows the team and individuals to understand the focus of their training. The simplest statistics are the number of occurrences or the frequency statistics. A frequency table is a method of organizing raw data in a compact form by displaying a series of scores in ascending or descending order, together with their frequencies. The number of times a given competency occurs across JDs can show the relative importance of the competencies.

5.3.3 Centrality measures – Degree centrality

Centrality is a simple method where we want to identify which nodes are in the 'center' of the network in the sense that they have many and important connections. Three standard centrality measures capture a wide range of 'importance' in a network. The centrality of a node measures its relative importance within the graph. They measure the amount of influence a node has over the information flowing between nodes. However, the removal of a centrally important node will destabilize an organization. There are various centrality measures as expanded in the next sections.

The degree centrality measures show how central a node is in a network, with higher degrees indicating greater importance. This centrality measure is based on the number of connections between the nodes. In contrast, the degree centrality of two nodes that are closely related will give a higher degree to the first. Degree centrality measures focus on immediate relationships and don't consider indirect connections. For instance, an actor might be closely linked to many neighbors, but isolated from a vast network. However, this doesn't mean that they are central to the network at large. The degree of a node is a summary of the number of neighbours the node has.

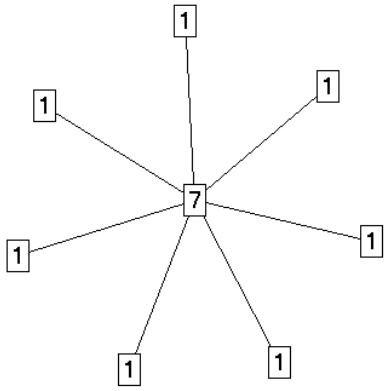


Figure 3: Centrality measure

Degree centrality is the simplest centrality measure to compute. The degree centrality for a node is simply its degree. A node with 10 social connections would have a degree centrality of 10. A node with 1 edge would have a degree centrality of 1. The nodes with higher outdegree is more central. The degree centrality of a node refers to the number of edges attached to the node. In order to know the standardized score, we need to divide each score by n-1 (n = the number of nodes). Since the graph has 7 nodes, 6 (7-1) is the denominator for this question. Here we divide the number of links directly connected to the node by n-1 where n means the total number of the nodes in focal network.

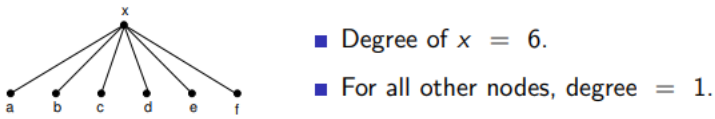


Figure 4: Example degree of centrality

5.3.4 Strength: Weighted Degree Centrality

Strength centrality is a measure that reflects the degree to which each node is connected to other nodes in the network. Weighted degree of strength considers the weight as well that represents the weight of the connections

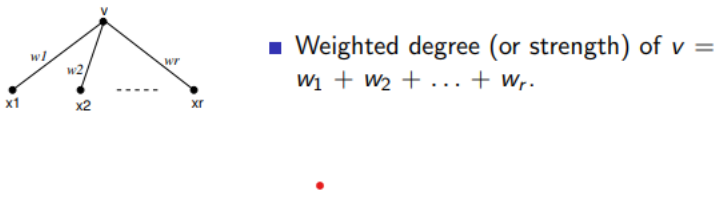


Figure 5: Example weighted degree of centrality

5.3.5 Gini impurity index (as the measure of distribution for labor market /as a whole

There are various methods to identify the importance of competencies. Neighbourhood metrics are derived from a node n-hop neighbourhood. Competencies have long been used as a framework to help focus employees' behavior on things that matter most to an organization and help drive success. It is extremely important for organizations to understand the competencies in descending order of importance. Competencies are presented in descending order of their importance. The Gini coefficient can serve to measure inequality in the importance of skills. Having eight competencies, the minimal value of Gini coefficient (that is 0) will be obtained when the contribution of every competency is the same and equal to $\frac{1}{8}$. On the other hand, the maximal value of its coefficient (that is 1), we will obtain if the significance of one competence is equal to 1, and for all other competencies their significance will be equal 0. Likewise, we can analyse the weights of a skill scheme.

For a set of samples X with k competencies.

$$gini(X) = 1 - \sum_{i=1}^{k} p_i^2$$

where pi is the proportion of competencies of class i

here is an example for the same:

Competencies	N	Pi
C1	3	3/17
C2	0	0
C3	4	4/17
C4	7	7/17
C5	2	2/17
C6	1	1/17
	17	

GI=1-0.27336=.72664

Table9: Gini Impurity Index example (Source Author)

5.3.6 Identification of groups of competencies:

The results obtained from the previous step (document-competency matrix) will be transformed into competency co-occurrence graphs that have a form of weighted graphs in which nodes represent competencies and the edges represent relationships between them. Weights are assigned to nodes (and they express competency significance) and to edges (they represent the strength of connections between competencies). It is worth underlining that competency co-occurrence graph can be built for an individual document that shows abilities of a given candidate or expectations of a given company or can be built from a set of documents and then it shows general regularities. For identification of the main competencies which are strongly related to each other, a process of identification of communities existing in a graph will be performed. Communities will be treated as patterns of related competencies (called competency schemas). Based on the internal and external information, competency schemes in the banking sector in the UAE would be analyzed.

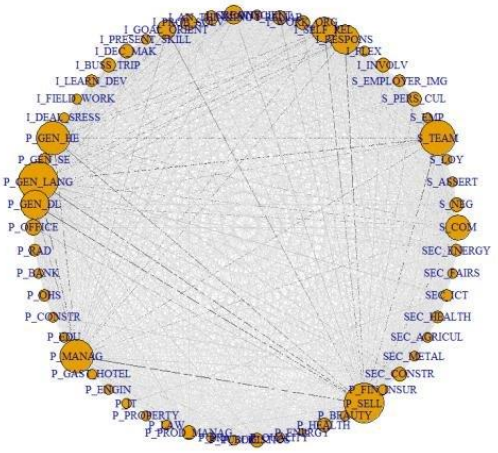


Figure6: Competency co-occurrence graph of Polish labor market (Source: Pawel Lula at. Al.)

5.3.6.1 Components (connected)

With the use of exploratory textual analysis techniques, seven hundred job vacancies were analysed, and the required competencies were identified. As part of the analysis, the system presented in Lula, P. et al. (2017) was used. The importance of competencies is represented by a diameter of nodes and the importance of relations by the darkness of edges. We first analysed the five categories of job competencies.

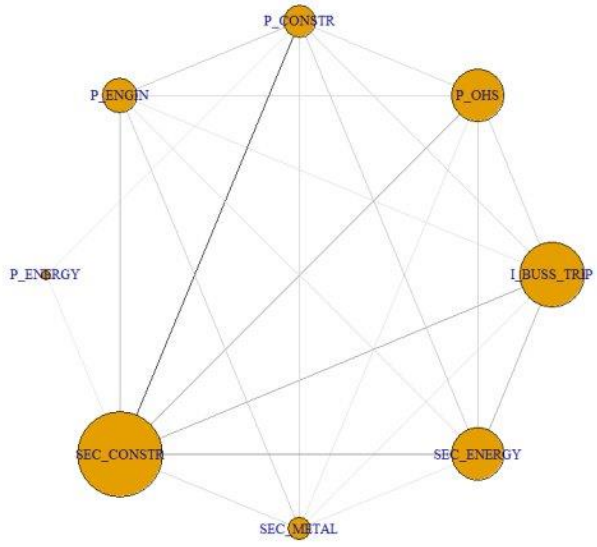


Figure 7: Competency schema identified on the Polish labor market (Source: Lula et al., 2019)

As per M.E.J.Neman (2006) ,Communities allow us to create a large scale map of a network since individual communities act like meta-nodes in the network which makes its study easier.

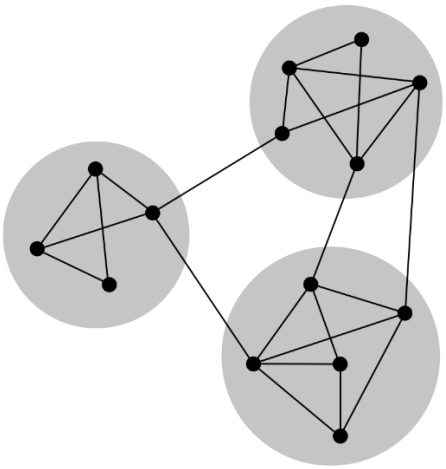


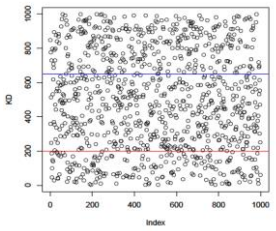
Figure: A sketch of a small network displaying community structure, with three groups of nodes with dense internal connections and sparser connections between groups. (source: J ham3)

5.3.7 K-Means segmentation

K-means is one of the most widely used unsupervised learning algorithms. It's also a powerful tool to identify unknown groups in a complex data set. However, it's not without its pitfalls. In some cases, the algorithm is ineffective. In other cases, it can create an enormous distance between data points. This is a problem if you have too many data points to fit into a cluster. K-means clustering works through an iterative process. It begins with the allocation of two centroids. Each centroids represents the average point in each cluster. In the next step, each data point is assigned to the centroid that is the closest to it. The algorithm repeats this process until the minimum change in the cluster centers is achieved. This process repeats several times until the best centroid initialization is determined. In addition, k-means is highly dependent on the initial values. In other words, if you're unsure of the value of k, the algorithm might not perform as well as one may like. This can be mitigated by running the algorithm with different initial values.

Steps for the K Means segmentation:

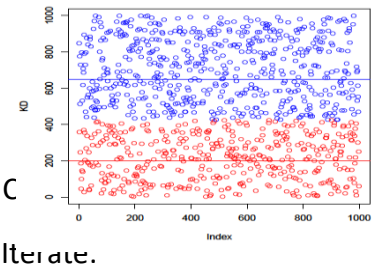
Basic Algorithm:

- Step 0: select K
- Step 1: randomly select initial cluster seeds
- 

- An initial cluster seed represents the “mean value” of its cluster.
- In the preceding figure:
    - Cluster seed 1 = 600
    - Cluster seed 2 = 200.

Step 2: calculate distance from each object to each cluster seed as squared Euclidean distance

Step 3: Assign each object to the closest cluster

Step 4: C  each cluster

Iterate:

- Calculate distance from objects to cluster centroids.
- Assign objects to closest cluster
- Recalculate new centroids

Stop based on convergence criteria

- No change in clusters
- Max iterations

Approach tries to minimize the within-cluster sum of squares error (WCSS)

- Implicit assumption that SSE is similar for each group

- The overall WCSS is given by: $\sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$
- The goal is to find the smallest WCSS.

5.4 Analysis of relationships between competences and other features

5.4.1 Bipartite models and statistics

Bipartite graphs consist of two sets of nodes. Each set has an edge connecting it. The two sets are usually referred to as top and bottom nodes. Bipartite graphs are also called bipartite models because of the way they are constructed. Each set has its own degree sequence, which may be different for the graphs. Bipartite graphs are often used to describe the behavior of networks. Bipartite networks are complex networks that contain two sets of nodes. The connections between these sets are only possible between elements in the two sets. Hence, the network can be modeled as a bipartite network with nodes in either set.

5.4.2 High level network-level statistics based on bipartite graphs and ecological models

This set of indices is computed for the entire network. The characteristics of network level statistics are presented in Dormann et al. (2009).

5.4.2.1 Specialization asymmetry:

In a study by Blüthgen, Menzel, and Blüthgen (2006), the specialization index, $H2$, is presented. To calculate the $H2$, the probabilities associated with the $\boldsymbol{G}$ matrix need to be defined. This can be done by summing up all the elements in the matrix.

The relations between elements of $\boldsymbol{C}$ and $\boldsymbol{V}$ sets are defined by an interaction matrix G:

$$G = \begin{matrix} g_{11} & .. & g_{1M} \\ .. & .. & .. \\ g_{N1} & .. & g_{NM} \end{matrix}$$

The $\boldsymbol{G}$ matrix columns represent competencies and rows – elements of the $\boldsymbol{V}$ set. Element gij gives the number of interactions between vi and cj values.

$$S = \sum_{i=1}^{v} \sum_{j-1}^{c} g_{ij}$$

Next, probabilities pij may be expressed as:

$p_{ij} = g_{ij} /s$

also, marginal probabilities can be calculated:

$$p_{i*} = \frac{\sum_{j=1}^{M} g_{ij}}{s}$$

and

$$p_{j*} = \frac{\sum_{i=1}^{N} g_{ij}}{s}$$

Having a probability matrix, the two-dimensional Shannon entropy can be expressed:

$$H_2 = \sum_{i=1}^{N} \sum_{j-1}^{M} (p_{ij} \ \ln p_{ij})$$

Lower H2 values are indicative of higher specialization and higher H2 values are indicative of higher generalization. Unfortunately, H2 values are not limited to the range [0; 1] and thus should be standardized.

5.4.2.2 Cluster coefficient

Cluster coefficient represents both the network-wide binary, one-mode-based cluster coefficient as well as those for each level. The idea of the cluster coefficient was introduced in Watts & Strogatz (1998). For a given node a cluster coefficient calculates the probability that selected at random, neighbours of u are connected by an edge (it means that they are neighbors to each other). In other words, it expresses the tendency to form a clique (a graph in which all nodes are connected directly by an edge) by neighbours of a given node u. The cluster coefficient calculation is relatively simple for one-mode networks. . Assume that two nodes are called

neighbours if an edge between them exists. Let $N(u)$ be a set of neighbours of a node u. Then a possible number of edges between neighbors is calculated:

$$t_{N(u)=} \frac{\mathrm{N(u)|(|N(u)|-1)}}{2}$$

where $|N(u)|$ is the number of neighbours of a node u.

5.4.2.3 Connectance:

Realized proportion of possible links (Dunne et al. 2002) is defined as the sum of links divided by the number of cells in the matrix (= number of higher times number of lower trophic level species). This is the standardized number of species combinations often used in co-occurrence analyses (Gotelli and Graves, 1996). The connectance index is defined as:

$$Wc = \frac{I}{M * N}$$

where I is the number of non-zero elements in the matrix $\boldsymbol{G}$. The Nc index takes into account only the existence of relations between elements of $\boldsymbol{C}$ and $\boldsymbol{V}$ sets and ignores its strength.

5.4.3 Chi-squared test and measures based on chi-squared statistics

The Chi-square test of independence checks whether two variables are likely to be related or not. We have counted for two categorical or nominal variables. The null hypothesis states that knowing the level of Variable A does not help you predict the level of Variable B. That is, the variables are independent. The alternative hypothesis is that knowing the level of Variable A can help you predict the level of Variable B.

5.4.3.1 Chi-squared test for goodness of fit

- State null hypothesized proportions for each category. The alternative is that at least one of the proportions is different than specified in the null.
- Calculate the expected counts for each cell as npi.
- Calculate the χ2 statistic:

$$\chi^2 = \sum \frac{(\text{observed - expected})^2}{\text{expected}}$$

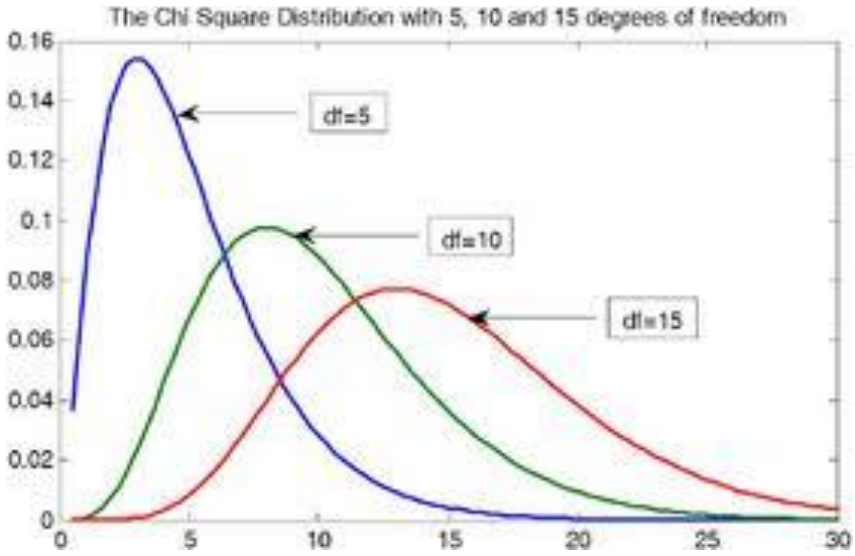


Figure8: An example key competency in descending order of importance (Source: Denise Jackson et al., 2012)

- Compute the p-value as the proportion above the χ2 statistic for either a randomization distribution or a χ2 distribution with df
- Interpret the p-value: "The exact p-value for a non-directional test is the sum of probabilities for the table having a test statistic greater than or equal to the value of the observed test statistic."
    - High p-value: High probability that test statistic > observed test statistic. Do not reject null hypothesis.
    - Low p-value: Low probability that test statistic > observed test statistic. Reject null hypothesis.

5.4.3.2 Cramer's V

It is a measure of association between two nominal variables, giving a value between 0 and +1 (inclusive). It is based on Pearson's chi-squared statistic and was published by Harald Cramér in 1946. These measures are appropriate when both variables are categorical — either nominal or ordinal—because neither assumes anything about the direction of a relationship.

$$\varphi = \sqrt{\frac{X^2}{(N)(df_{row/column})}}$$

Cramer's V is based on χ^2 and takes a value between 0, no relationship, and 1, a perfect relationship.

5.5 Summary and conclusion

In this chapter we explained in the general description of models used for analysis and prediction of labor market followed by the explanation of graph-based models used for analysis of the labor market.

Chapter 6: Model of employers' expectations towards candidates for employment in banking sector in the UAE

6.1 Data retrieving and process of model construction

6.1.1 General information and scope

The standalone competency assessment may not be an appropriate and it is worth investigating the relationship among them while analyzing the job descriptions for industry. In general, modern companies are interested in four group of the core competencies such as intellectual, professional, personal, and interpersonal (Cichoń M., Piotrowska I., 2018). Here, we extend the same and have a group of the core competencies such as Academic, Industry, personal, technical and workforce related to the current study.

6.1.2 Scope

In ontology-based approach, we are going to browse all Job descriptions and find all phrases and key words matching of patterns defined within the ontology and competency schemas defined in chapter 5. As a result, a matrix will be created with rows related to documents and columns related to competencies. Elements of this matrix will represent a contribution of a given competency in each document. The results obtained from the previous step (document-competency matrix) will be transformed into competency co-occurrence graphs which have a form of weighted graphs in which nodes represent competencies and the edges represent relationships between them. Weights are assigned to nodes (and they express competency significance) and to edges (they represent the strength of connections between competencies). For identification of the main competencies which are strongly related to each other, a process of identification of communities existing in a graph will be performed. Communities will be treated as patterns of related competencies (called competency schemas). We wish to perform this for high level and detailed competencies schemas. Also, we wish to test if there is any significant job posting difference between the major cities in the UAE.

6.1.3 Demography: Regions, Cities:

Within UAE, Dubai and Abu Dhabi are the two major cities where most of the population stays. While performing the analysis, we have also created the demographic variable for geography. All the Job descriptions are divided in three groups. Dubai, Abu Dhabi and others.

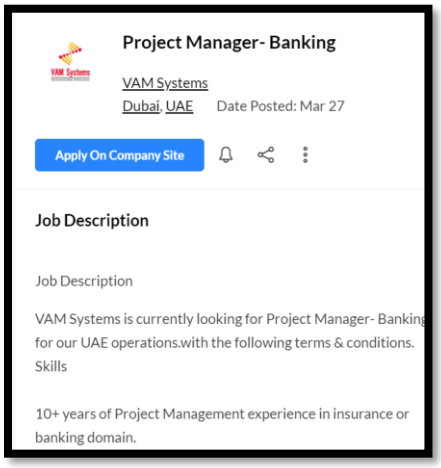


Figure 1: A sample Job description from banking sector in UAE (Source: Bayt.com)

6.1.4 Data extraction

The Job descriptions have been used for pre and post covid for the same. We have extracted the Job Descriptions from gulf talent, Bayt and Naukri gulf as explained in chapter 4 and details are provided there.

6.1.4.1 Web scraping:

Web scraping is a method of gathering large amounts of data from the internet, often referring to the process of crawling websites for information. The process is analogous to knocking on a door and requesting permission to access a specific site. Once the site has granted permission, web scraping tools send an HTTP request to gather data about that site. The data is then used on a website for visitors to search. A web scraper accesses XML and HTML files from websites and parses them to produce structured data. The data is stored locally, making it available for analysis. Web scrapers access XML or HTML content, and then parse the data, usually storing it locally. They can extract a predefined set of data, and they can be easily customized to meet the needs of their users. Here we have created a web scraper to extract

data from three job sites for our analytical purposes. As the structure of the website is different from all the three job sites, we had to customize and create three different web scrappers.

6.1.4.1 Matrix Structure:

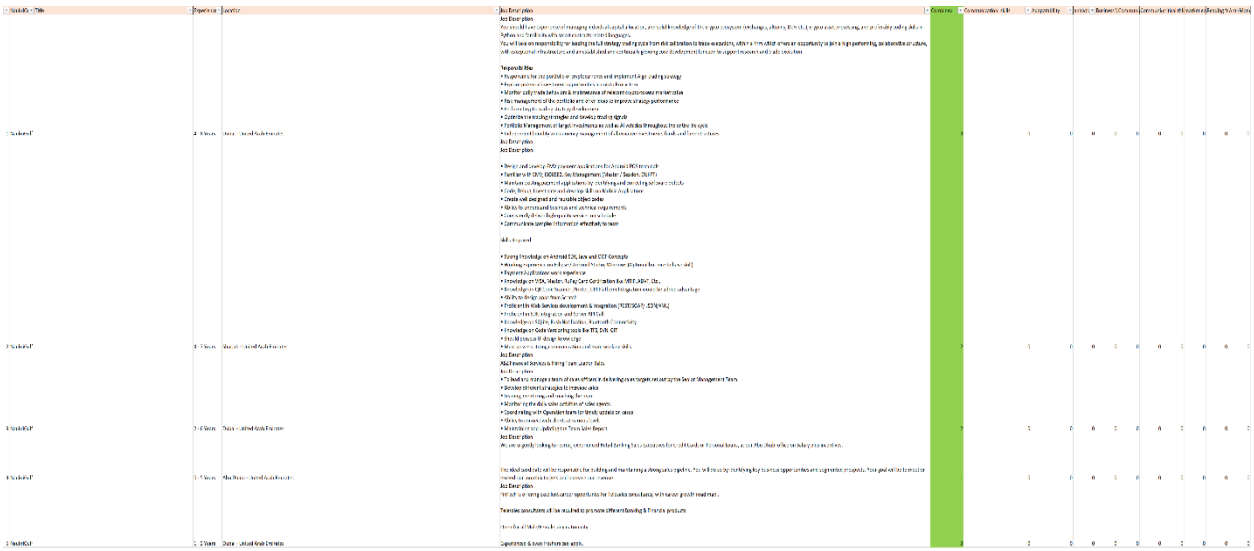


Table 1: Industry wide competencies (Source: Author)

6.1.5 Detailed information about competencies which were identified

We have proposed an 84-competency (detailed) model for the banking industry in UAE. This framework can help employees understand the skills that they must possess. While academic excellence is an obvious prerequisite for jobs in the banking sector, other skills such as industry knowledge are often a prerequisite. These additional skills are often called a competency model and are closely associated with specific job roles. The accelerated pace of digitalization and pandemic economic situation have compelled employers to consider a focus on special skills and competencies in the banking industry. In a competency model for banking, employees need to be capable of thinking beyond their immediate area of expertise. The most effective candidates can frame questions and solve problems across disciplines. They should understand systems and technology. They should also be able to leverage diverse perspectives and build a valuable

network of contacts. This new approach is changing the way banks do business. This shift in the way they operate is paving the way for new growth and profit opportunities.

In today's world, companies have moved toward remote working styles. Virtual meetings and conferences are increasingly common. Automated systems can generate advice, but a human touch is necessary to make people act on it. A competency model for banking could also serve as the point of contact between an automated system and the mass market. By developing a flexible content management system that integrates the various data sources, banks can create a rich library of content and improve their customer profiles.

Core Competencies
Academic Competencies
Industry wide competencies
Personal Effectiveness
Technical Competencies
workplace competencies

Table2: proposed high level competency categories for the banking sector in UAE

6.2 Analysis of competencies:

The role of competences and their relationships are crucial for understanding the fitment of profile for the job. It is extremely important to understand and identify which competencies matter the most. Here, we focus on their interpretations typically focus on core competencies, such as reading, writing, math, and science. However, the importance of analyzing the interplay between competencies and their relationships cannot be overemphasized. There are several reasons for this. One possible explanation is that age is negatively related to cognitive capacities and competencies. Older cohorts are often disadvantaged in their learning, as their specific educational experiences differ from those of the younger cohorts. Accordingly, the authors of the competency taxonomy tended to overlook the taxonomy, which is also a factor in evaluating the competency. Nevertheless, these two factors are closely related. Regardless of the

motivations for defining and categorizing competencies, they provide guidance for settling interpretive differences.

Another explanation is that some of the skills that are common in higher education are also associated with low levels of literacy in each country. In the US, for example, most college graduates were under-educated. In contrast, a large minority of people in low-education countries were literate. This could reflect the structure of the labor market in a particular country. Thus, the relationship between formal education and competencies is affected by compositional effects.

6.2.1 Detailed Competency co-occurrence table:

competency co-occurrence table allows individuals or organizations to identify the relationships and interdependencies between different competencies. The table can be used to identify which competencies are often used together, which ones are needed to achieve certain goals or tasks, and which ones may be lacking in a particular team or organization.

A detailed competency co-occurrence table has been created like below. It is not to show the whole matrix so maybe the form like:

	Analytical_thinking	Anti_Money_Laundering	Audit	Branch_administration
Analytical_thinking	3	0	0	1
Anti_Money_Laundering	0	5	0	2
Audit	0	0	5	0
Branch_administration	1	2	0	105
Business_strategy	0	0	0	0
Business_Writing	0	1	0	1
Client_advisory	0	1	1	5
Content_writer	0	0	0	0
Critical_thinking	0	0	0	2
Financial_reporting	0	0	0	0
Fintech	0	0	0	2
Flexibility	0	0	0	0
Fraud_preventions	1	1	1	3
health_and_safety	0	0	0	0
Information_Technology	0	0	0	5
Insurance	0	0	0	11
Integrity	0	0	0	0
Interpersonal_Skills	2	3	2	3
Leadership_Skills_	2	2	1	41
Legal	0	0	0	0
legal_and_compliance	1	2	1	7
Management_information	0	0	0	5
market_economics	0	1	0	1
marketing	0	0	0	17
Organizational_developme	0	0	0	10
Privatization_restructuring	2	1	2	22
Procurement_manager	3	3	2	34
project_management	2	2	0	8
Reading_for_information	0	0	0	0
Regulations	2	0	0	17
Reliability	0	0	1	0
Risk_Management	0	0	0	0
Self_Management	0	0	0	0
Teamwork	0	0	0	0
Time_Management	0	0	0	0
transportation_manager	0	2	1	5
warehouse_manager	1	0	4	25
Wealth_Management	0	0	1	1
tax	0	0	1	19
Communication_listening	0	0	0	18
Communication_speaking	0	0	0	0
Customer_service	0	0	1	16
Software	0	0	0	0
Ethics	1	1	0	18
Big_Data_Analysis	0	0	0	1
Programming	0	3	2	10
UI_Design	0	0	0	3
Social_Media_Managemer	0	0	0	1
Microsoft_Office	0	0	0	1
Cloud	0	0	0	0

Table3: Detailed Competency co-occurrence table (sample) Source: Author

In this table, each competency is listed in the first row and column. The percentages in the table represent the degree to which each competency is related to the other competencies listed. The table can be used to identify areas where additional competencies may be needed. Overall, a competency co-occurrence table can be a useful tool for individuals and organizations to identify areas of strength and weakness in their competencies and to develop strategies for improvement.

6.2.2 Importance of competencies:

How to show importance of competency is extremely important to understand the key competency in demand. With a globalization of consumer behaviour and an increased focus on digital transformation, many banks are overly focused on IT transformation and business innovation. Regulatory guidance has driven the creation of new and innovative products and services based on customer journeys. Hence, competency demand is changing more than before, hence we need a framework to understand key competencies in demand.

6.2.2.1 number of occurrences/Frequency table:

A frequency table is a method of organizing raw data in a compact form by displaying a series of scores in ascending or descending order, together with their frequencies. Here we can have a look at the number of occurrences for key competencies from the below table.

Here is the frequency table for the same at the high level competencies:

Competencies	Frequecies of occurrences
Sum of Academic Competencies	7
Sum of Industry wide competencies	494
Sum of Personal Effectiveness	133
Sum of Technical Competencies	60
Sum of Workplace Competencies	219

Table4: Frequency table (Source: Author)

Now if we look at the detailed competencies,

Competencies	Frequecies of occurences
Customer service	174
Procurement manager	157
Leadership Skills	148
Regulations	139
Branch administration	105
warehouse manager	75
Privatization, restructuring, acquisition and merger	66
Ethics	65
Fraud preventions	55
legal and compliance	47
marketing	43
Communication listening	41
Programming	33
project management	31
transportation manager	25
Organizational development and human resources management	24
Interpersonal Skills	20
Management information systems	18
Fintech	17
Insurance	16
Reliability	16
Critical thinking	15
Integrity	15
Client advisory	13
Information Technology	10
Big Data Analysis	8
market economics	7
Communication speaking	7
Anti-Money Laundering	5
Audit	5
Business Writing	5
health and safety	4
Wealth Management	4
Analytical thinking	3
Business strategy	3
Financial reporting	3
UI Design	3

Risk Management	2
Content writer	1
Flexibility	1
Legal	1
Reading for information	1
Self Management	1
Teamwork	1
Time Management	1
Software	1
Social Media Management	1
Microsoft Office	1

Table 5: number of occurrences of key competencies (Source: Author)

Such ranking needs to be generated more frequently implementing these solutions, a bank will ensure its continuity of operations and remain competitive. However, it is also essential for a bank to focus on its core competencies, as this will help it maintain its strong position in the industry.

6.2.2.2 Heat Map:

We can also have a look at the heat map to understand the relationship as well.

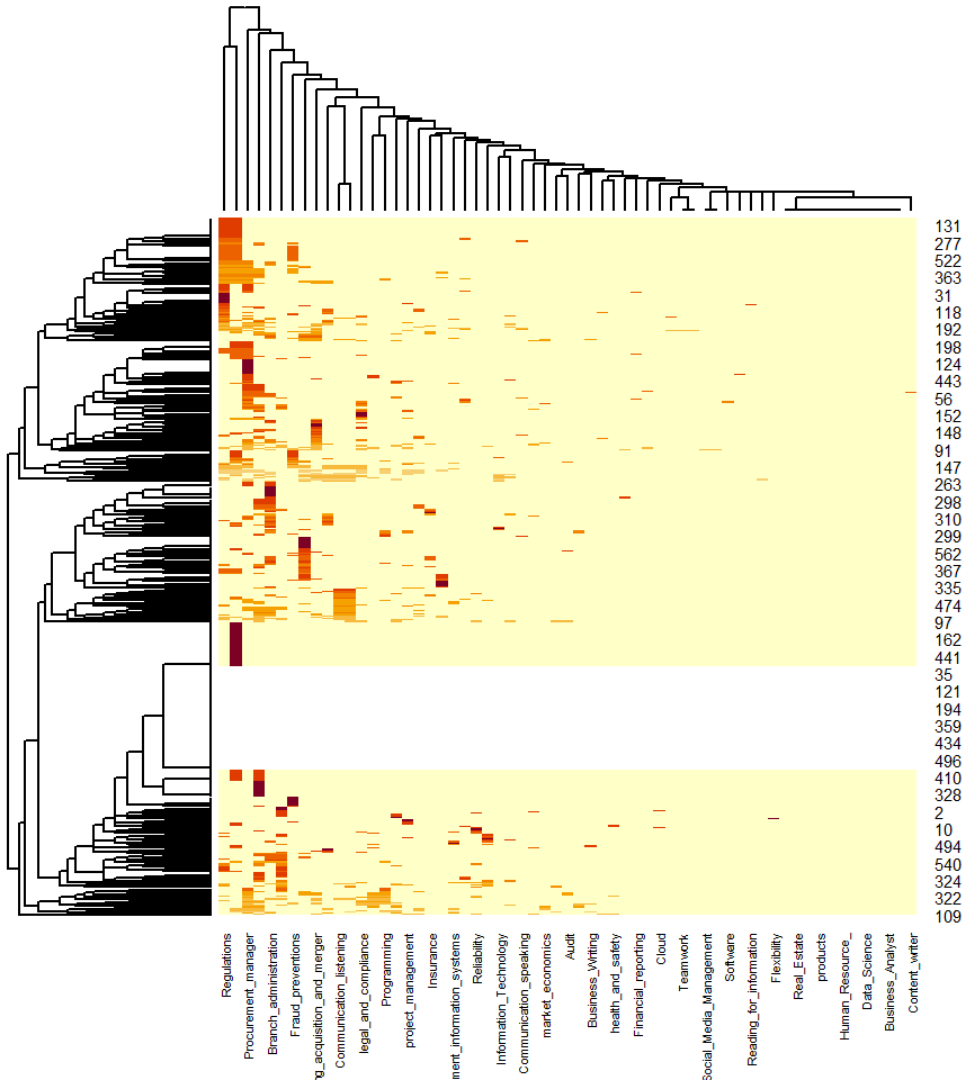


Figure 4: Heatmap (Source: Author)

6.2.3 Competency co-occurrence graph:

There are many approaches to define employee competencies which differentiate competencies, one of the best approaches is a competency co-occurrence graph. A co-occurrence graph is a powerful tool for identifying patterns in text. They can also help us understand the similarities and differences between two competencies. The terms that co-occur are grouped into neighborhoods based on their interconnections.

Individual terms may have many neighbors or be connected through more than one term. However, the neighborhoods are not connected to each other.

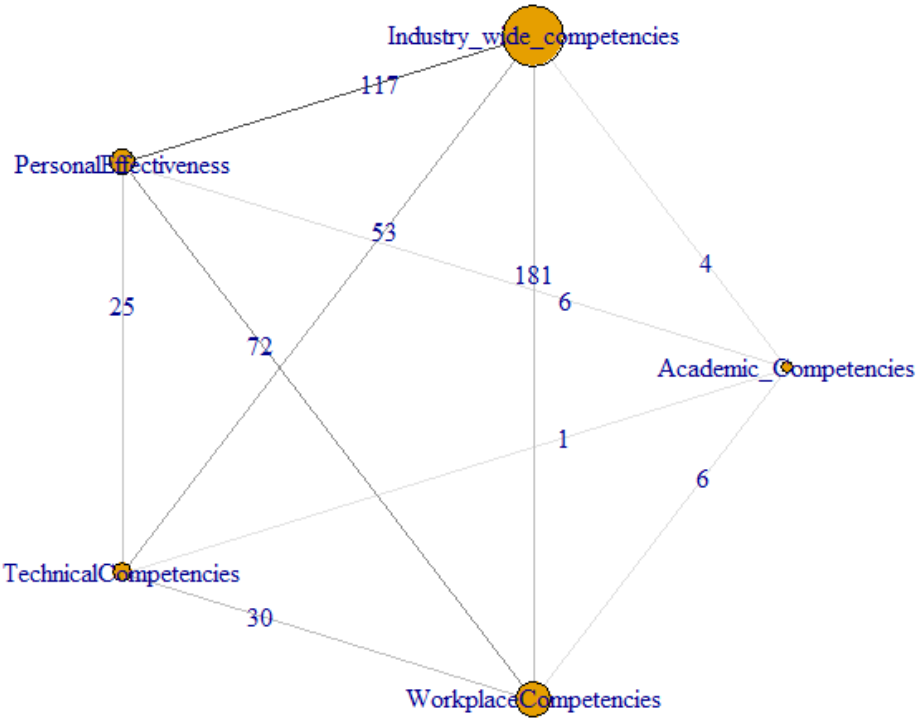


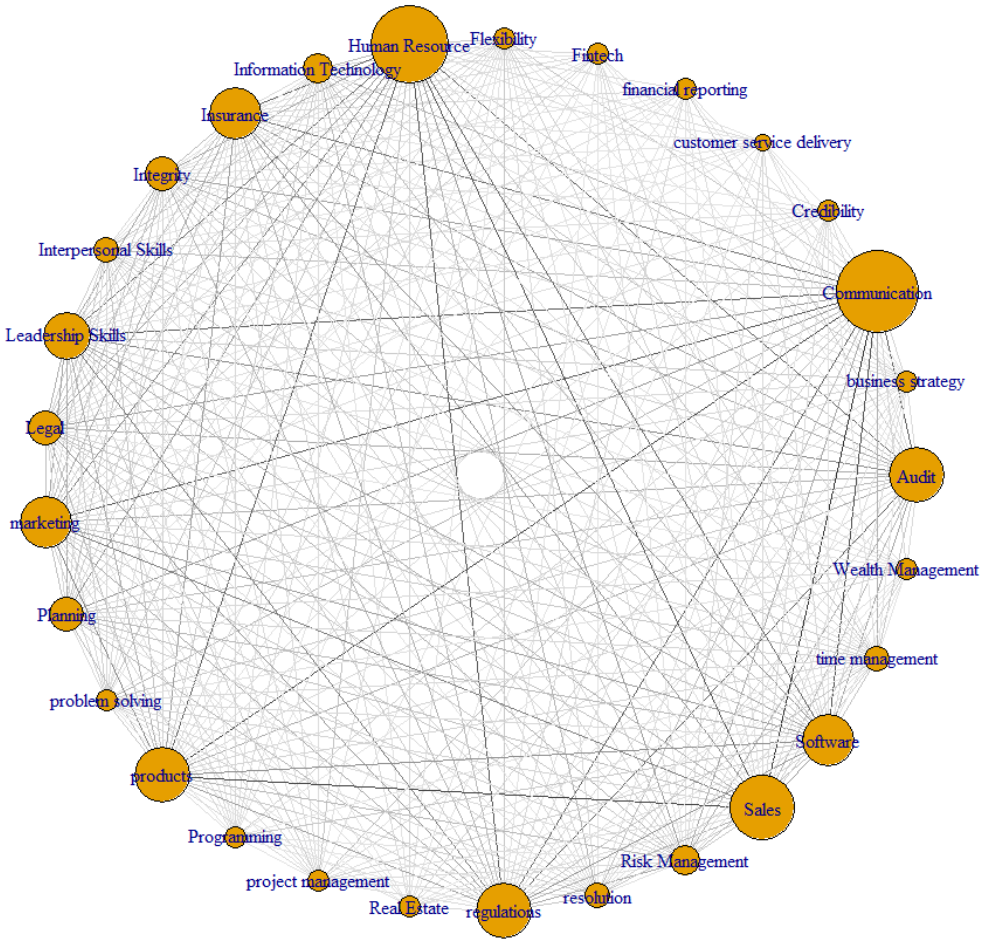
Figure2: Competency co-occurrence graph of UAE Banking sector labor market at a category level (Source: Author)

Here, we provide the competency co-occurrence graph that tells how the high-level competencies are connected to each other. Also, the darkness of the line connecting the competencies represents the frequency of the relationship. For an example personal competency along with industry wide competencies appears 117 times together.

These networks are a network of connections between two high level competencies. Typically, the co-occurrence graph can help us explain how different competencies are Co-related. Here, we find pairs of terms that occur together. These terms are called neighbors. These neighbors are connected through a single term. Neighborhoods may consist of many terms. In some cases, a single term has several neighbors. It can be noticed that for the competency schema presented in Figure 1, the contribution of two the most significant competencies are the industry and personal effectiveness. We find the industry competencies playing the significant role. In our preliminary experiments and results, we determined the industry skills are the key requirements in the banking sector in UAE, followed by personal and workplace competencies.

6.2.4 Detailed competency co-occurrence graph:

As a next step, we expanded the Competency co-occurrence using detailed competency schema. Here we can find the top 29 detailed competencies and how they are connected to each other. Our detailed competency schemas have total 84 competencies. There are many competencies that are not well connected and do not appear here.



Figgure3: Detailed Competency co-occurrence graph of UAE Banking sector labor market at a category level (Source: Author)

We can see that sales, knowledge of banking products, audit, communication and marketing competencies are the largest one as per table 2.

6.2.5 Centrality measures:

One of the ways to measure the relative importance of different skills is by using the centrality measures. This method scores for nodes by their 'closeness' to other nodes and assigns a score to each based on the length of each link between them. Closeness centrality is useful for determining which nodes are influential in a specific cluster. This metric is useful for evaluating the relative influence of a particular skill, such as

sales or audit. A centrality measure is a measure of how important a node is within a network. It counts the number of times each node lies on the shortest path between two nodes. This measure can be useful for analyzing the communication dynamics in a network. If a node has a high Betweenness value, it implies that the node has considerable authority over a disparate cluster. The degree centrality measure quantifies the degree of centrality of a node to its neighbors. It measures the proportion of links that pass through a node. The higher the number, the more important that the node is to the network. A degree of centrality indicates how central a particular node is to a network. Similarly, the percolation centrality measure is a quantitative measure of how important a node is to its neighbors. The degree centrality of a node is the distance between a node and all other nodes. The distance between nodes depends on whether the node is connected to more than one node. In undirected graphs, the distance between two nodes is irrelevant. However, it can influence the results of centrality. A website may have high closeness centrality when connected to incoming and outgoing links. Another centrality measure is the indulge of a node. Indegree is the number of ties that a node has two other nodes. Positive in degrees indicates a node's popularity and gregariousness. Inverses centrality can also indicate a node's influence on a network. An indegree is the number of ties that bind a node to others.

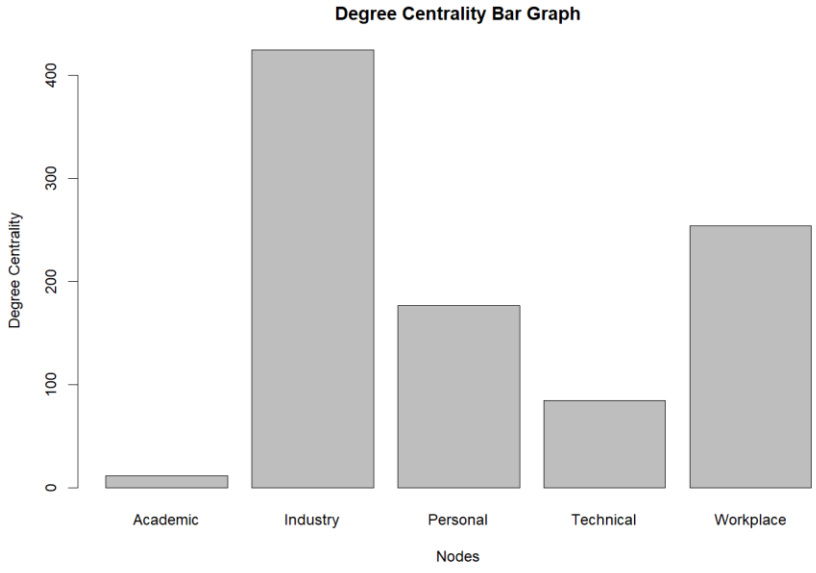


Figure 4: degree centrality bar graph for high level competencies

6.2.5.1 High level competency and its importance using Gini Index.

The calculations of Gini coefficient as performed for the whole market (using frequencies for every competency). The calculations of Gini coefficient should be performed for the whole market (using frequencies for every competency Competencies are presented in descending order of their importance. The column “Cum. w.” (Cumulative weights) shows for every competency the meaning of a given competency and all preceding ones) from table 3. The Gini coefficient can serve to measure inequality in the importance of skills. Having eight competencies, the minimal value of Gini coefficient (that is 0) will be obtained when the contribution of every competency is the same and equal to $\frac{1}{8}$. On the other hand, the maximal value of its coefficient (that is 1) we will obtain if the significance of one competence is equal to 1, and for all other competencies their significance will be equal 0. Likewise, we can analyse the weights of a skill scheme.

Node	Weight	Cum. w.
"Ethics	0.233	0.233"
"legal_and_compliance	0.165	0.398"
"Programming	0.113	0.511"
"project_management	0.105	0.617"
"transportation_manager	0.083	0.699"
"Interpersonal_Skills	0.068	0.767"
"Critical_thinking	0.053	0.82"
"Reliability	0.053	0.872"
"market_economics	0.023	0.895"
"Big_Data_Analysis	0.023	0.917"
"Anti_Money_Laundering	0.015	0.932"
"Audit	0.015	0.947"
"Business_Writing	0.015	0.962"
"Analytical_thinking	0.008	0.97"
"Business_strategy	0.008	0.977"
"health_and_safety	0.008	0.985"
"Wealth_Management	0.008	0.992"

"Cloud	0.008	1"

Table 6: The importance of key high-level competencies (Source: Author)

We can see the ethics, Legal and compliance, programming, big data analytics, AML regulations and leadership skills are the most important one.

6.2.5.2 Detailed competency and its importance:

Here we have the detailed competency edge that are significant in the descending order of importance.

Edge	Weight	Cum. w.
Audit|regulations	0.0706	0.0706
Leadership Skills|Regulations	0.0567	0.1273
Audit|Leadership Skills	0.0481	0.1754
Legal|regulations	0.0417	0.2171
Integrity|Leadership Skills	0.0332	0.2503
Audit|Risk Management	0.031	0.2813
regulations|Risk Management	0.031	0.3123
Audit|Legal	0.0299	0.3422
Leadership Skills|Planning	0.0299	0.3722
Leadership Skills|Legal	0.0289	0.4011
Audit|Integrity	0.0278	0.4289
Leadership Skills|Risk Management	0.0267	0.4556
Integrity|regulations	0.0257	0.4813
Planning|regulations	0.0235	0.5048
Information Technology|regulations	0.0225	0.5273
Audit|Planning	0.0214	0.5487
Audit|Information Technology	0.0182	0.5668
regulations|time management	0.0182	0.585

Integrity|Planning	0.0171	0.6021
Audit|financial reporting	0.015	0.6171
Flexibility|regulations	0.015	0.6321
Legal|Risk Management	0.015	0.6471
Audit|time management	0.015	0.662
business strategy|Leadership Skills	0.0139	0.6759
Integrity|Legal	0.0139	0.6898
Leadership Skills|project management	0.0139	0.7037
Information Technology|Risk Management	0.0139	0.7176
Leadership Skills|time management	0.0139	0.7316
Information Technology|Leadership Skills	0.0118	0.7433
Integrity|Risk Management	0.0118	0.7551
Planning|Risk Management	0.0118	0.7668
Flexibility|Leadership Skills	0.0107	0.7775
Flexibility|time management	0.0107	0.7882
Flexibility|Legal	0.0086	0.7968
Information Technology|Planning	0.0086	0.8053
Audit|project management	0.0086	0.8139
financial reporting|regulations	0.0086	0.8225
Legal|time management	0.0086	0.831
Planning|time management	0.0086	0.8396
business strategy|Integrity	0.0075	0.8471
Information Technology|Legal	0.0075	0.8545
Legal|Planning	0.0075	0.862
Planning|project management	0.0075	0.8695
Information Technology|time management	0.0075	0.877
Flexibility|Information Technology	0.0064	0.8834

business strategy|Planning	0.0064	0.8898
Flexibility|Planning	0.0064	0.8963
business strategy|regulations	0.0064	0.9027
Risk Management|time management	0.0064	0.9091
Audit|business strategy	0.0053	0.9144
business strategy|Legal	0.0053	0.9198
financial reporting|Legal	0.0053	0.9251
Integrity|project management	0.0053	0.9305
project management|regulations	0.0053	0.9358
project management|Risk Management	0.0053	0.9412
Flexibility|Integrity	0.0043	0.9455
Information Technology|project management	0.0043	0.9497
Legal|project management	0.0043	0.954
Flexibility|Risk Management	0.0043	0.9583
Audit|Flexibility	0.0032	0.9615
business strategy|Flexibility	0.0032	0.9647
business strategy|Information Technology	0.0032	0.9679
Information Technology|Integrity	0.0032	0.9711
financial reporting|Leadership Skills	0.0032	0.9743
business strategy|project management	0.0032	0.9775
Flexibility|project management	0.0032	0.9807
business strategy|Risk Management	0.0032	0.984
Integrity|time management	0.0032	0.9872
project management|time management	0.0032	0.9904
business strategy|financial reporting	0.0021	0.9925
financial reporting|Integrity	0.0021	0.9947
business strategy|time management	0.0021	0.9968

financial reporting|Planning	0.0011	0.9979
financial reporting|project management	0.0011	0.9989
financial reporting|Risk Management	0.0011	1
financial reporting|project management	0.0011	0.9989
financial reporting|Risk Management	0.0011	1

Table 7: The distribution of weights in the competency schema (Source: Author)

6.3 bipartite network indices: Compare competencies network with cities and without cities of job posting

The bipartite network characterization framework has remained essentially unchanged for a long time. Indexes are a common tool to analyze networks and can be used to compare different configurations of networks. They are insensitive to species interactions with one another, and they may not fully capture complex relationships. These indices can also lead to substantial loss of information, as they may not be appropriate for understanding the role of individual species in a network. In qualitative networks, bipartite motifs are defined as the number of species that interact with one another. In quantitative networks, interactions are weighted in proportion to their relative strength, which can be misleading because rare species can have a disproportionate influence on network metrics. Moreover, conventional indices are less granular when it comes to indirect interactions, which are typically hidden in the topology. Bipartite network indices are calculated using a topology function. It is important to note that network indices have been derived from data pertaining to different pollinator groups and are therefore likely to have different applications. Some indices are simple descriptions of the network's topography, while others are more complex.

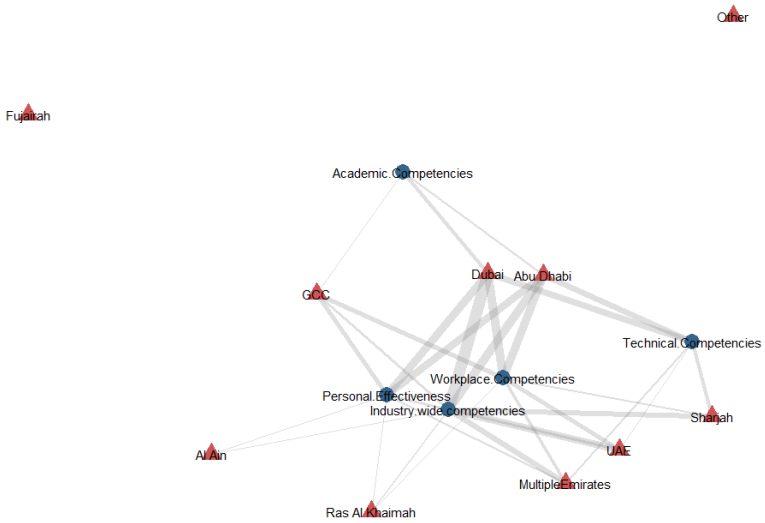


Figure 8: Bipartite graph for city of job posting and key high-level competencies

6.3.1 Compare competencies network with cities and without cities of job posting: Using high level competency schema

We wish to understand here if city of job posting does have an impact or not. So, we are building two networks here, one with city of job posting and another without it. Next, we calculate the network statistics for both and find the difference in network, we calculated the below set of indices is computed for the entire network level. They are computed twice. Once without the city of posting and once with city of postings. It is worth noticing that network-wide specialization index is significant and show that the competency needed for the cities is different in UAE for the banking sector.

Network stats	
5 Key Competencies without City of Postings	**5 Key Competencies with City of Postings**
H2	H2
"0"	"0.0846572799890386"
specialization asymmetry	specialization asymmetry
"0.218856990554771"	"0.733561016297872"

cluster coefficient	cluster coefficient
"0.369140625"	"0.421378091872792"
connectance	connectance
"0.38671875"	"0.458186101295642"

Table 8: Network statistics for High level competencies without and with city of Job Postings (Source: Author)

Connectome i.e. The e standardized number of species combinations often used in co-occurrence analyses, or the clustering coefficient that is the mean, across all species, of the number of realized links divided by the number of possible links for each species (i.e. Average per-species connection) is very different between the two networks.

6.3.2 Compare competencies network with cities and without cities of job posting: Using detailed competency schema

Network statistics for the detailed competencies with and without city of job postings has been computed and as per below table.

Network stats	
All key competencies without City of Postings	**All Key Competencies with City of Postings**
H2	H2
"0"	"0.0551830302734726"
specialisation asymmetry	specialisation asymmetry
"0.352666159573296"	"0.49072618669666"
cluster coefficient	cluster coefficient
"0.0880626223091976"	"0.0901060070671378"
connectance	connectance

"0.142182333490789"	"0.157420494699647"

Table 9: Network statistics for detailed competencies without and with city of Job Postings (Source: Author)

Connectance i.e., the e standardized number of species combinations often used in co-occurrence analyses, or the clustering coefficient that is the mean, across all species, of the number of realized links divided by the number of possible links for each species (i.e. Average per-species connection) is very different between the two networks.

6.4 Test of the independence of competencies across cities:

The Chi-square test of independence verifies whether two variables are likely to be linked or not. It (also known as the Pearson Chi-square test) is one of the most useful statistics for testing hypotheses when the variables are nominal. The Cramer's V is the most common strength test used to test the data when a significant Chi-square result has been obtained. It is interpreted as a measure of the relative (strength) of an association between two variables. The coefficient ranges from 0 to 1 (perfect association). In practice, an Cramer's V of. 10 provides a good minimum threshold for suggesting there is a substantive relationship between two variables.

6.4.1 high Level Competencies

Statistics for Industry wide competencies by Location			
Statistic	**DF**	**Value**	**Probe**
Chi-Square	3	13.946	0.003

Table 8: Chi Square results for high level competencies and locations (Source: Author)

As per the results and Cramer V, we find that the null hypothesis is rejected for the industry competency. We found that different industry competencies are required in different cities in United Arab Emirates.

6.4.2 Detailed Competencies

We also performed the Chi Square test on the detailed competencies to check if the competencies do differ by location. We have found that the competencies like risk management, sales, regulatory competencies, and others differ across cities as per below table.

Statistics in Table of Credibility by Location				Statistics in Table of workplace competencies by Location			
Statistic	**DF**	**Value**	**Probe**	**Statistic**	**DF**	**Value**	**Prob**
Chi-Square	3	14.92	0.002	**Chi-Square**	3	15.98	0.001
Human Resource management by Location				**Information Technology by Location**			
Statistic	**DF**	**Value**	**Prob**	**Statistic**	**DF**	**Value**	**Probe**
Chi-Square	3	14.831	0.002	**Chi-Square**	3	8.5573	0.036
Integrity by Location by Location				**Table of Legal by Location**			
Statistic	**DF**	**Value**	**Prob**	**Statistic**	**DF**	**Value**	**Prob**
Chi-Square	3	14.192	0.003	**Chi-Square**	3	8.3348	0.04

Table of Legal by Location				Table of Products Knowledge by Location			
Statistic	**DF**	**Value**	**Probe**	**Statistic**	**DF**	**Value**	**Prob**
Chi-Square	3	7.1452	0.067	**Chi-Square**	3	16.986	7.00E-04
Table of Real Estate by Location				**Table of Regulations by Location**			
Statistic	**DF**	**Value**	**Probe**	**Statistic**	**DF**	**Value**	**Prob**
Chi-Square	3	7.8987	0.048	**Chi-Square**	3	10.916	0.012
Table of Risk Management by Location				**Sales by Location**			
Statistic	**DF**	**Value**	**Probe**	**Statistic**	**DF**	**Value**	**Prob**
Chi-Square	3	5.8956	0.117	**Chi-Square**	3	12.231	0.007
time management by Location							
Statistic	**DF**	**Value**	**Prob**				
Chi-Square	3	15.206	0.002				

Table 10: Chi Square results for detailed competences and locations (Source: Author)

This table only lists the competency that is found significantly in the chi - square test. 4.5 Conclusion and Discussions:

The Chi-square test of independence checks whether two variables are likely to be related or not. We can see from table 6 and 7 that there are various competencies that are in demand but differ in various cities.

6.5 Segmentation of JDs based on key competencies at a high level:

K-means clustering aims to partition Job Description data into k clusters in a way that data points in the same cluster are similar and data points in the different clusters are farther apart. The similarity of two points is determined by the distance between them. Here, we perform segmentation using the 5 Key high-level competencies to create clusters of JDs.

6.5.1 Cubic Clustering Criterion to decide on the number of clusters:

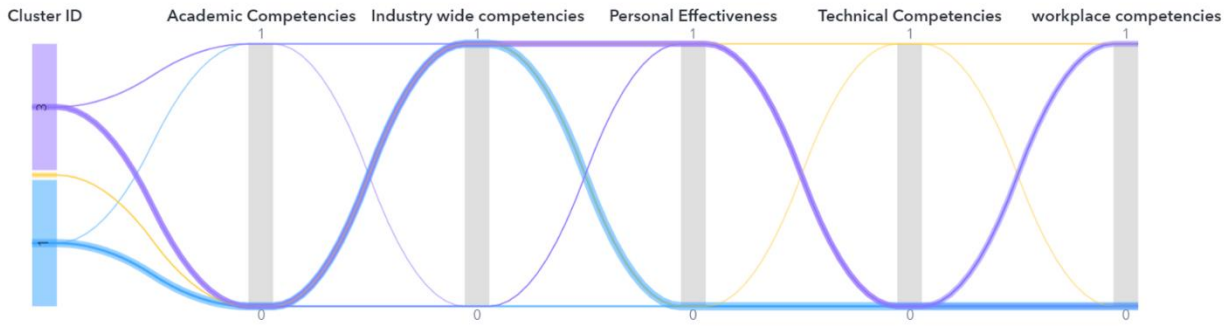
The cubic clustering criterion (CCC) has been used to estimate the number of clusters using k - means based on minimizing the within-cluster sum of squares. Based on the CCC and minimum 30 records per cluster, we resulted in 3 clusters as described below.

6.5.2 Cluster summary:

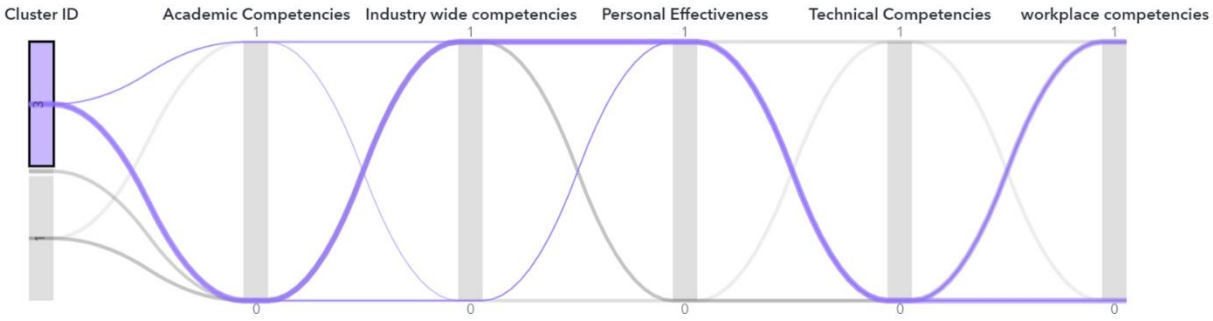
Cluster Summary						
Cluster	**Frequency**	**RMS Std Deviation**	**Maximum Distance from Seed to Observation**	**Radius Exceeded**	**Nearest Cluster**	**Distance Between Cluster Centroids**
1	219	0.3198	1.4736		3	1.0375
2	30	0.2682	1.0734		3	1.0419
3	407	0.2525	1.2278		1	1.0375

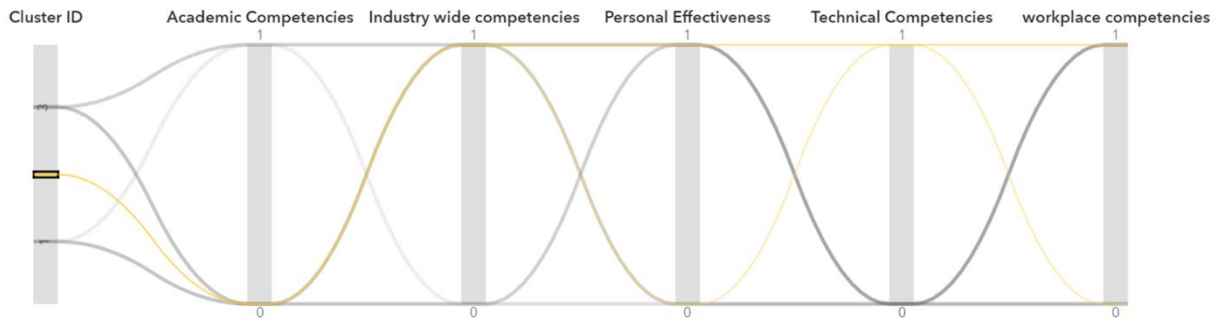
Table 11: cluster statistics (Source: Author)

6.5.3 Cluster profiling:



A parallel coordinate plot maps each row in the data table as a line, or profile. Each attribute of a row is represented by a point on the line. This makes parallel coordinate plots similar in appearance to line charts, but the way data is translated into a plot is substantially different. Parallel coordinate graph for the three clusters are as below:





And the cluster 3

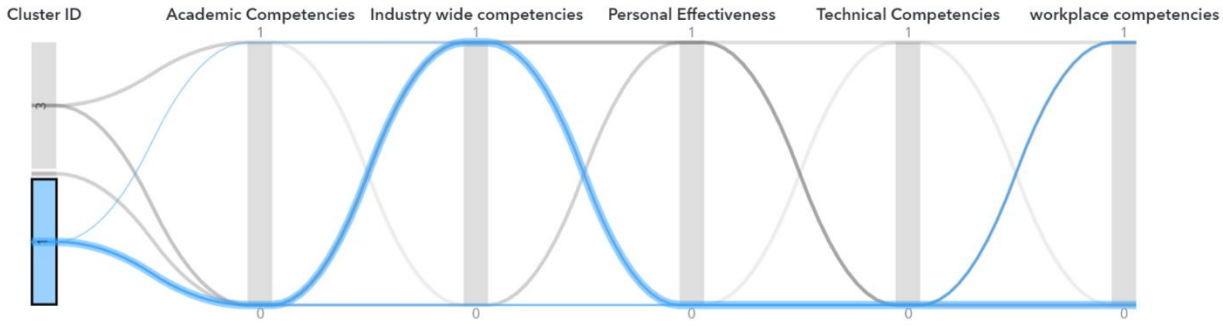


Figure 9: Parallel co ordinates and clustering profiling (source: Author)

A different segment of job descriptions can help identify the clusters of JDs and how they are influenced by the competency in demand. This helps us understand the heterogeneity that exists in the Job description Data.

6.6 Summary and conclusions:

In this chapter, the focus was on extracting data, building models, and analyzing competencies and their relationships. The skills analysis helped to identify crucial high-level competencies for the banking sector in the United Arab Emirates. It was found that industry skills play a significant role and are the primary requirements, followed by personal and professional skills. The skill co-occurrence chart showed that both industry and individual skills are necessary, as well as workplace and individual skills. These results emphasized the importance of both joint and individual skills in the UAE banking sector.

The results showed that ethics, legislation and compliance, programs, big data analysis, AML regulations, and leadership competencies are the most important skills in the UAE banking sector. The detailed skills analysis also revealed key relationships and their importance, listed in decreasing order. The study demonstrates that the competence scheme concept can be applied to represent crucial models in the labor market, providing valuable insight into the requirements of employers. This information can be used to assess competency gaps, design studies or training programs tailored to the needs of the labor market. The authors used the R program to perform the calculations and the Chi Square Test of Independence to identify the skills in demand that vary between different cities.

In this chapter, the focus was on analyzing the employment opportunities in the banking sector of the UAE. Different methods for competency analysis were presented, with an emphasis on examining the relationships between competencies and other factors such as positions and locations. Understanding these relationships is crucial to determine if the demand for competencies varies in different cities.

7. Chapter 7: Recommendations and further research

The author's dissertation focuses on analyzing the expectations of employers in the Banking Sector in the UAE regarding the competencies of job candidates using a Text Mining Approach. The author recognized the need for a scientific competency framework in the banking industry in the UAE and proposed a framework accordingly. To study the representation of the crucial competencies in the domain knowledge, the author conducted an ontology/text mining-based study, examining the views of employers, employees, and job seekers in the banking sector in the UAE.

7.1 Conclusions:

Our study first analyzed the public and private banking sectors and the role of the banking sector in the economy of the UAE to gain a deeper understanding of the country. The banking sector review includes a detailed examination of the structure of both public and private banks in the UAE. Additionally, the chapter presents relevant labor market statistics.

To test the hypothesis 1 that states that the exploratory text analysis of job descriptions can be used for identification of crucial competencies expected in the banking sector in the UAE, research reviewed previous studies on the analysis of competencies across various industries, with a focus on the banking sector. To perform a job effectively, organizations must evaluate the required competencies, identify any gaps, and establish a strategy to address them. To validate the hypothesis, the theoretical foundation, and methodological considerations for using text mining to analyze the Job discrepancy database is performed. The text mining approach was thoroughly explained, including the use of unstructured data visualization and topic modeling. Data was collected from public job sites in the banking sector and analyzed using corpus-based methods and corpus-based text mining to identify key competencies and their relationships. The analysis was conducted using the Latent Dirichlet Allocation method, a probabilistic model that assumes each document contains a mixture of topics. The key findings after term clustering of job descriptions have been performed. This analysis results in various focus areas of hiring in the

banking sector. There are two key findings from this analysis Front-office was made redundant faster (direct sales agents) or through outsourced parties, now they are needed after such a hockey stick recovery. Also, back-office has been made bare thin and out- sourced and investments reduced. This made back-office and middle-office staff not being required much during fill-back period. Risk management has been a growing category where there has been a focus in hiring in banking sector in the UAE. Competencies demanded by companies are changing faster now than before (Mckinsey 2020). AI systems mainly NLP, can be used to identify the emerging trends in the hiring by industry. Governments, University, and private institutes must play an active role in creating awareness of emerging technology trends. This unique application of topic modeling can be used for further representation of crucial patterns, changes in the labor market in the banking sector.

A second hypothesis revolved around building a competency schema that allows for an in-depth description and examination of relationships between expected competencies in the banking sector in UAE. The study implemented an ontology-based approach to analyze the competencies expected in the banking sector in the UAE using a proposed schema. The ontology-based approach uses labeled graph-based models to describe a given domain, with the objects represented by nodes and relationships between objects described by edges. This approach was used to analyze published job descriptions from the banking sector to identify key competencies and their relationships. The results showed that the competency scheme can effectively represent the crucial schemes in the labor market in the UAE. The scheme provides insight into the requirements of employers and can be useful for assessing competency gaps, designing training programs, and conducting further studies. The proposed approach offers the advantage of incorporating context information into the competency analysis, allowing for the estimation of the importance and specificity of competencies and context factors, and the analysis of relationships between them. The results presented in the paper show that the concept of the skill system can be used to represent the crucial models existing in the labor markets. The

schemes can provide a detailed insight into employers' needs. It can also help estimate the skills gap or design education or training programs adapted to labor market needs.

Last and the third hypothesis states that the network models are useful for description of various aspects of labor market in banking sector in the UAE. We presented the theoretical underpinnings and introduced graph-based models for text mining job dependencies. Graphs are used to represent the co-occurrence of words in text segments or documents, making them valuable for identifying important keywords and key phrases. According to the study, industry skills are the most critical requirement for success, followed by personal and professional skills. The skill co-occurrence chart highlighted the significance of both industry and individual skills in this sector. Therefore, the findings stress the importance of both joint and individual competencies for a thriving career in the UAE banking industry. The research revealed that ethics, legislation and compliance, programs, big data analysis, AML regulations, and leadership competencies are the most essential skills in the UAE banking industry. The study employed the competence scheme concept to represent significant models in the labor market, providing valuable insight into the requirements of employers. The authors utilized the R program and the Chi Square Test of Independence to identify sought-after skills that differ across various cities. The research aimed to analyze employment prospects in the UAE banking sector while examining the associations between competencies, positions, and locations. Understanding these connections is crucial in identifying whether the demand for skills fluctuates in the distinct cities. This information can help to evaluate competency gaps and create targeted studies or training programs that align with the needs of the labor market. In the post-COVID era, organizations are rapidly changing their product portfolios, technologies, and customer experience strategies. This is reflected in the fast-changing job description data extracted from the banking sector in the UAE. The thesis makes a significant contribution by identifying the key competencies in demand in the banking sector. Hence, we could build a network model based on bipartite graphs explaining relationships between various aspects of labor market in UAE.

7.2 Recommendations and Future Research:

Here we provide details for limitations or shortcomings of the study and about directions for further research based on the results of the study. Also, recommendations from thesis for practitioners and policy makers are provided below.

7.2.1 Recommendations for practitioners:

This a unique competency model to perform the analysis of competency gap in the UAE banking sector. This helps in identifying crucial competencies expected in the banking sector in the UAE along with identifying competencies possessed by candidates looking for employment. We build a software tool for performing a complete analysis of the supply side and demand side of this labor market that can be used for any industry and market.

7.2.2 Recommendations for Future research:

Future research contributing to this model needs to focus on getting more data for some other cities that are not so well represented in the data as there were very limited JDs posted for the cities. The issues on competency gap in various labor market, decision making on changeable requirements to competencies of jobseekers etc. Also in cross-country perspective should be studied in further research. Topic modeling is a widely used approach in the field of natural language processing for discovering the latent topics presented in a collection of text documents. There are several other well-known topic modeling algorithms such as Non-Negative Matrix Factorization (NMF) that can be used to extend empirical research on topic modeling. There are several avenues that can be explored as well.

Model improvements: we can try to improve existing topic modeling algorithms by proposing modifications to the models, such as incorporating prior knowledge into the models or adding regularization terms to the objective function.

Model evaluation: There is a need for more robust evaluation metrics for topic models. We can work on developing new evaluation metrics that better capture the quality of the topics generated by the models.

Domain adaptation: Topic models can be adapted to specific domains by incorporating domain-specific knowledge into the models. We can explore ways to incorporate domain-specific information, such as domain-specific stop words or domain-specific priors, into the models to improve their performance.

Model interpretability: Topic models are often criticized for their lack of interpretability. We can work on developing techniques for improving the interpretability of topic models, such as visualization techniques or techniques for generating meaningful topic labels.

Overall, there are many directions in which empirical research on topic modeling can be extended. By exploring these avenues, we can work towards developing more effective and interpretable topic models for a wide range of applications.

7.2.3 Recommendations for policymakers:

As this model can help identify the skills gaps and the competencies needed by the industry, private and public policy makers can encourage continuous professional development (CPD) practices using this information. The banking industry should encourage and incentivize employees to participate in continuous professional development programs. This can be done by providing financial support, flexible work schedules, and recognition for completing courses or certifications. This tool can also help foster corporate trainings as companies can foster a culture of learning by encouraging employees to share knowledge and skills with each other. This can be done through mentorship programs, cross-functional training, and knowledge-sharing events. Invest in technology and innovation: Banks can invest in technology and innovation to stay competitive and improve efficiency. This includes providing training on new technologies, encouraging experimentation and innovation, and providing opportunities for employees to work on new projects.

8. List of bibliography

Ademiluyi, L. F. (2019). Employability skills needed by business education graduates as perceived by business teachers and employers of labour in two Southwestern Nigerian States. Business Education Innovation Journal, 11(1), 57-65..

Agazzi, E. (2011). Consistency, truth and ontology. Studia Logica, 97, 7-29

Aggarwal, C. C. (2018). Machine learning for text (Vol. 848). Cham: Springer.

Alalwan, J., & Thomas, M. (2012). An ontology-based approach to assessing records management systems. e-Service Journal: A Journal of Electronic Services in the Public and Private Sectors, 8(3), 24-41.

Ali, Y., & Salih, M. (2016). Stakeholder Competencies Intelligence-Scale Development and Validation Some Evidence from KSA. Journal of Leadership, Accountability & Ethics, 13(1)..

Alghamdi, R., & Alfalqi, K. (2015). A survey of topic modeling in text mining. Int. J. Adv. Comput. Sci. Appl.(IJACSA), 6(1).

Al-Obeidat, F., Kafeza, E., & Spencer, B. (2018). Opinions sandbox: turning emotions on topics into actionable analytics. In Emerging Technologies for Developing Countries: First International EAI Conference, AFRICATEK 2017, Marrakech, Morocco, March 27-28, 2017 Proceedings 1st (pp. 110-119). Springer International Publishing.

Aparicio, M., Bacao, F., & Oliveira, T. (2016). An e-learning theoretical framework. An e-learning theoretical framework, (1), 292-307.

Antonelli, D., Bruno, G., Taurino, T., & Villa, A. (2015). Graph-based models to classify effective collaboration in SME networks. International Journal of Production Research, 53(20), 6198-6209.
Osei, A. J., & Ackah, O. (2015). Employee's competency and organizational performance in the pharmaceutical industry. International Journal of Economics, Commerce and Management, 3(3), 1-9.

Arslan, A., & Yener, S. (2020). I like my leader; not yours!. Transylvanian Review of Administrative Sciences, 16(59), 5-22.

Asghar, Z., Ali, T., Ahmad, I., Tharanidharan, S., Nazar, S. K. A., & Kamal, S. (2019). Sentiment analysis on automobile brands using Twitter data. In Intelligent Technologies and Applications: First International Conference, INTAP 2018, Bahawalpur, Pakistan, October 23-25, 2018, Revised Selected Papers 1 (pp. 76-85). Springer Singapore.

Askar, P., & Altun, A. (2009). CogSkillnet: An ontology-based representation of cognitive skills.

Asmussen, C. B., & Møller, C. (2019). Smart literature review: a practical topic modelling approach to exploratory literature review. Journal of Big Data, 6(1), 1-18.

Atkociuniene, Z. 2010, 'Knowledge management information in improving the organization's competencies, Information sciences', Vol. 21, No.5, pp. 52 - 57.

Baker, T., McKay, I., Morden, D. L., Dunning, K., & Schuster, F. E. (1996). Breakthrough in organization performance: Competitive advantage through employee-centered management. People and Strategy, 19(4), 14.

Banazir, B., & Philip, A. (2013). Efficient Keyword Search Using Text Mining Techniques: A Survey. International Journal of Engineering and Innovative Technology (IJEIT), 9001(1), 2277-3754.

Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of management, 17(1), 99-120.

Bandura, A., Freeman, W. H., & Lightsey, R. (1999). Self-efficacy: The exercise of control.

Barber, M. J. (2007). Modularity and community detection in bipartite networks. Physical Review E, 76(6), 066102.

Barker, N., Davis, C. A., López-Peña, P., Mitchell, H., Mobarak, A. M., Naguib, K., ... & Vernot, C. (2020). Migration and the labour market impacts of COVID-19 (No. 2020/139). WIDER Working Paper.

Bartik, A. W., Bertrand, M., Lin, F., Rothstein, J., & Unrath, M. (2020). Measuring the labor market at the onset of the COVID-19 crisis (No. w27613). National Bureau of Economic Research.

Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of management, 17(1), 99-120.

Beckett, S. J. (2016). Improved community detection in weighted bipartite networks. Royal Society open science, 3(1), 140536.

Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2016). Paper recommender systems: a literature survey. International Journal on Digital Libraries, 17, 305-338.

Bell, D. N., & Blanchflower, D. G. (2020). US and UK labour markets before and during the Covid-19 crash. National Institute Economic Review, 252, R52-R69.

Benedetto, F., & Tedeschi, A. (2016). Big data sentiment analysis for brand monitoring in social media streams by cloud computing. Sentiment Analysis and Ontology Engineering: An Environment of Computational Intelligence, 341-377.

Berry, M. W., Dumais, S. T., & O'Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. SIAM review, 37(4), 573-595.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Bikse, V., Lusena-Ezera, I., Rivza, P., & Rivza, B. (2021). The development of digital transformation and relevant competencies for employees in the context of the impact of the COVID-19 pandemic in latvia. Sustainability, 13(16), 9233.

Borland, J., & Charlton, A. (2020). The Australian labour market and the early impact of COVID-19: An assessment. Australian Economic Review, 53(3), 297-324.

Boyatzis, R. E. (1991). The competent manager: A model for effective performance. John Wiley & Sons.

Boyatzis, R. E., Stubbs, E. C., & Taylor, S. N. (2002). Learning cognitive and emotional intelligence competencies through graduate management education. Academy of Management Learning & Education, 1(2), 150-162.

Boussiakou, L. G., Boussiakou, I. K., & Kalkani, E. C. (2006). Student development using emotional intelligence. World Transactions on Engineering and Technology Education, 5(1), 53.

B Božina Beroš, M. (2018). Some reflections on the governance framework of the single resolution board. JCMS: Journal of Common Market Studies, 56(3), 646-655.

Cao, N., & Cui, W. (2016). Introduction to text visualization.

Caruth, J., Middlebrook, B. & Rachel, F. (1985): Overcoming Resistance to Change, Advanced Journal of Management, Vol. 50 Iss. 3 P.23

Carson, K. D., & Carson, P. P. (1998). Career commitment, competencies, and citizenship. Journal of Career Assessment, 6(2), 195-208.

Chakraborty, G., Pagolu, M., & Garla, S. (2014). Text mining and analysis: practical methods, examples, and case studies using SAS. SAS Institute.

Chakraborty, G., Pagolu, M., & Garla, S. (2013). Text mining and analysis: Practical methods. Examples, and case studies using SAS, 181-184.

Chand, P. K., Kumar, A. S., & Mittal, A. (2019). Emotional intelligence and its relationship to employability skills and employer satisfaction with fresh engineering graduates. International Journal for Quality Research, 13(3), 735.

Chaudhary, S., & Singh, S. (2016). A Study of Competency Mapping of Employees in Banking Sector (With Special Reference to ICICI Bank). International Journal of Human Resources Management (IJHRM), 5(3), 11-20.

Chen, H. (2004). An intelligent broker architecture for pervasive context-aware systems. PhD Dissertation, University of Maryland

Chen, W. K., Chen, L. S., & Pan, Y. T. (2021). A text mining-based framework to discover the important factors in text reviews for predicting the views of live streaming. Applied Soft Computing, 111, 107704.

Cheng, X., Cao, Q., & Liao, S. S. (2022). An overview of literature on COVID-19, MERS and SARS: Using text mining and latent Dirichlet allocation. Journal of Information Science, 48(3), 304-320.

Choi, D., & Kim, P. (2013). Sentiment analysis for tracking breaking events: a case study on twitter. In Intelligent Information and Database Systems: 5th Asian Conference, ACIIDS 2013, Kuala Lumpur, Malaysia, March 18-20, 2013, Proceedings, Part II 5 (pp. 285-294). Springer Berlin Heidelberg.

Chang, J., & Blei, D. M. (2010). Hierarchical relational models for document networks.

Chiesa, R., Van der Heijden, B. I., Mazzetti, G., Mariani, M. G., & Guglielmi, D. (2020). "It is all in the game!": The role of political skill for perceived employability enhancement. Journal of Career Development, 47(4), 394-407.

Cichoń, M., & Piotrowska, I. (2018). Level of academic and didactic competencies among students as a measure to evaluate geographical education and preparation of students for the demands of the modern labour market. Quaestiones Geographicae, 37(1), 73-86.

Cohen, A. M., & Hersh, W. R. (2005). A survey of current work in biomedical text mining. Briefings in bioinformatics, 6(1), 57-71.

Dainty, A. R., Cheng, M. I., & Moore, D. R. (2004). A competency-based performance model for construction project managers. Construction Management and Economics, 22(8), 877-886.

David, M. E., David, F. R., & David, F. R. (2021). Closing the gap between graduates' skills and employers' requirements: a focus on the strategic management capstone business course. Administrative sciences, 11(1), 10.

De Bel-Air, F. (2015). Demography, Migration, and the Labour Market in the UAE.

Delucchi, K. L. (1983). The use and misuse of chi-square: Lewis and Burke revisited. Psychological Bulletin, 94(1), 166.

De Vos, A., De Hauw, S., & Van der Heijden, B. I. (2011). Competency development and career success: The mediating role of employability. Journal of vocational behavior, 79(2), 438-447.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American society for information science, 41(6), 391-407.

Dinkić, N., Džaković, N., Joković, J., Stoimenov, L., & Đukić, A. (2018). Using sentiment analysis of Twitter data for determining popularity of city locations. In ICT innovations 2016: cognitive functions and next generation ICT systems (pp. 156-164). Springer International Publishing.

Dischinger, J. (2006). The emerging supply chain management profession. Supply chain management review, v. 10, no. 1 (Jan./Feb. 2006), p. 62-68: ill.

Dorado, R., & Ratté, S. (2016, March). Semisupervised text classification using unsupervised topic information. In The Twenty-Ninth International Flairs Conference.

Dubois, D. D. (1993). Competency-based performance improvement: A strategy for organizational change. HRD Press, Inc., 22 Amherst Road, Amherst, MA 01002.

Drucker, P. (2014). Innovation and entrepreneurship. Routledge.

Dunbar, K., Laing, G., & Wynder, M. (2016). A Content Analysis of Accounting Job Advertisements: Skill Requirements for Graduates. E-Journal of Business Education and Scholarship of Teaching, 10(1), 58-72.

Dwivedi, D. N., Mahanty, G., & Vemareddy, A. (2022). How Responsible Is AI?: Identification of Key Public Concerns Using Sentiment Analysis and Topic Modeling. International Journal of Information Retrieval Research (IJIRR), 12(1), 1-14.

Dwivedi, D. N., & Anand, A. (2021). The Text Mining of Public Policy Documents in Response to COVID-19: A Comparison of the United Arab Emirates and the Kingdom of Saudi Arabia. Zarządzanie Publiczne/Public Governance, (1 (55)), 8-22.

Dwivedi, D. N., & Pathak, S. (2022). Sentiment analysis for COVID vaccinations using Twitter: text clustering of positive and negative sentiments. In Decision Sciences for COVID-19: Learning Through Case Studies (pp. 195-203). Cham: Springer International Publishing.

Dwivedi, D. N., & Anand, A. (2022). A comparative study of key themes of scientific research post COVID-19 in the United Arab Emirates and WHO using text mining approach. In Advances in Data and Information Sciences: Proceedings of ICDIS 2021 (pp. 341-350). Singapore: Springer Singapore.

Dwivedi, D., & Vemareddy, A. (2023, January). Sentiment Analytics for Crypto Pre and Post Covid: Topic Modeling. In Distributed Computing and Intelligent Technology: 19th International Conference, ICDCIT 2023, Bhubaneswar, India, January 18–22, 2023, Proceedings (pp. 303-315). Cham: Springer Nature Switzerland.

Dwivedi, D. N., Wójcik, K., & Vemareddyb, A. (2022). Identification of key concerns and sentiments towards data quality and data strategy challenges using sentiment analysis and topic

modeling. In Modern Classification and Data Analysis: Methodology and Applications to Micro- and Macroeconomic Problems (pp. 19-29). Cham: Springer International Publishing.

Dwivedi, D. N., Mahanty, G., & Vemareddy, A. (2022, October). Sentiment Analysis and Topic Modeling for Identifying Key Public Concerns of Water Quality/Issues. In Proceedings of the 5th International Conference on Water Resources (ICWR)–Volume 1: Current Research in Water Resources, Coastal and Environment (pp. 341-355). Singapore: Springer Nature Singapore.

Eirinaki, M., & Vazirgiannis, M. (2003). Web mining for web personalization. ACM Transactions on Internet Technology (TOIT), 3(1), 1-27.

Elrehail, H., Harazneh, I., Abuhjeeleh, M., Alzghoul, A., Alnajdawi, S., & Ibrahim, H. M. H. (2019). Employee satisfaction, human resource management practices and competitive advantage: The case of Northern Cyprus. European Journal of Management and Business Economics, 29(2), 125-149.

Eskici, H. B., & Koçak, N. A. (2018). A text mining application on monthly price developments reports. Central bank review, 18(2), 51-60.

Esiyok, C., & Albayrak, S. (2015). Twitter sentiment tracking for predicting marketing trends. Smart Information Systems: Computational Intelligence for Real-Life Applications, 47-74.

Etzioni, O., Banko, M., Soderland, S., & Weld, D. S. (2008). Open information extraction from the web. Communications of the ACM, 51(12), 68-74.

Fabritius, C. V., Madsen, N. L., Clausen, J., & Larsen, J. (2006). Finding the best visualization of an ontology. Journal of the Operational Research Society, 57, 1482-1490.

Fischer, A., Keller, A., Frinken, V., & Bunke, H. (2010, August). HMM-based word spotting in handwritten documents using subword models. In 2010 20th International Conference on Pattern Recognition (pp. 3416-3419). IEEE.

Fernández-Huerga, E. (2019). The labour demand of firms: an alternative conception based on the capabilities approach. Cambridge Journal of Economics, 43(1), 37-60.

Flamholtz, E. G., & Lacey, J. (1981). The implications of the economic theory of human capital for personnel management. Personnel Review, 10(1), 30-40.

Firat, A., Madnick, S., & Grosof, B. (2004). Contextual alignment of ontologies for semantic interoperability. Available at SSRN 612472.

Foltz, P. W. (1990). Using latent semantic indexing for information filtering. ACM SIGOIS Bulletin, 11(2-3), 40-47.

Fortunato, S. (2010). Community detection in graphs. Physics reports, 486(3-5), 75-174.

Furnas, G. (1988). Using latent semantic analysis to improve information retrieval. In Proceedings of the ACM Conference on Human Factors in Computing Systems (pp. 281-285). ACM Press.

Gallon, M. R., Stillman, H. M., & Coates, D. (1995). Putting core competency thinking into practice. Research-Technology Management, 38(3), 20-28.

Ganesh, A. (2012). A study of training and employee development in commercial. Journal of Commerce & Accounting Research, 1(2).

García, R., & Gil, R. (2008). A Web Ontology for Copyright Contract Management. International Journal of Electronic Commerce, 12(4), 99-114.

Giber, D., Lam, S. M., Goldsmith, M., & Bourke, J. (Eds.). (2009). Linkage Inc's best practices in leadership development handbook: Case studies, instruments, training. John Wiley & Sons.

GoleGolec, A., & Kahya, E. (2007). A fuzzy model for competency-based employee evaluation and selection. Computers & Industrial Engineering, 52(1), 143-161.

Gomez, S. J., & Peter, A. (2017). Developing a framework for employability skills of management graduates ,International Journal of Research in Commerce & Management, 8(10).

Gordon, M. D., & Dumais, S. (1998). Using latent semantic indexing for literature based discovery. Journal of the American Society for Information Science, 49(8), 674-685.

Gottipati, S., Shankararaman, V., & Lin, J. R. (2018). Text analytics approach to extract course improvement suggestions from students' feedback. Research and Practice in Technology Enhanced Learning, 13, 1-19.

Gross Domestic Product. (2018). 1–6. https://doi.org/10.1787/g2d71aef-en

Gottipati, S., Shankararaman, V., & Lin, J. R. (2018). Text analytics approach to extract course improvement suggestions from students' feedback. Research and Practice in Technology Enhanced Learning, 13, 1-19.

Greenstine, A. J. (2017). Diverging ways: On the trajectories of ontology in Parmenides, Aristotle, and Deleuze. Contemporary encounters with ancient metaphysics, 202.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge acquisition, 5(2), 199-220.

Godea, A. K., Caragea, C., Bulgarov, F. A., & Ramisetty-Mikler, S. (2015). An analysis of twitter data on e-cigarette sentiments and promotion. In Artificial Intelligence in Medicine: 15th Conference on Artificial Intelligence in Medicine, AIME 2015, Pavia, Italy, June 17-20, 2015. Proceedings 15 (pp. 205-215). Springer International Publishing.

Gupta, A., Dwivedi, D. N., Shah, J., & Saroj, R. (2021). Understanding Consumer Product Sentiments through Supervised Models on Cloud: Pre and Post COVID. Webology, 18(1), 406-415.

Guerrero, O. A., & Axtell, R. L. (2013). Employment growth through labor flow networks. PloS one, 8(5), e60808..

Guillén Ramo, L., Saris, W. E., & Boyatzis, R. E. (2009). The impact of social and emotional competencies on effectiveness of Spanish executives. Journal of Management development, 28(9), 771-793.

Harvey, L. (1997). The Student Satisfaction Manual (Buckingham, Open University Press/Society for Research into Higher Education).

Harvey, L., Harvey, L., Locke, W., & Morey, A. (2002). Enhancing employability, recognising diversity: Making links between higher education and the world of work: Main report. Universities UK.

H Harmsen, H., & Jensen, B. (2004). Identifying the determinants of value creation in the market: A competence-based approach. Journal of Business Research, 57(5), 533-547.

Heijde, C. M. V. D., & Van Der Heijden, B. I. (2006). A competence-based and multidimensional operationalization and measurement of employability. Human Resource Management:

Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management, 45(3), 449-476.

Hippner, H., & Rentzmann, R. (2006). Text mining. Informatik-Spektrum, 29, 287-290.

Hirschhorn, L. (1984). Beyond mechanization (pp. 41-47). Cambridge, ma: mit Press.

Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. Machine learning, 42(1-2), 177.

Horrocks, I. (2008). Ontologies and databases. Presentation at the Semantic Days, 4.

Hoskisson, R. E., Gambeta, E., Green, C. D., & Li, T. X. (2018). Is my firm-specific investment protected? Overcoming the stakeholder investment dilemma in the resource-based view. Academy of Management Review, 43(2), 284-306.

IDC 2021,Worldwide Global DataSphere and Global StorageSphere Structured and Unstructured Data Forecast, 2021–2025

Jackson, S. E., Schuler, R. S., & Jiang, K. (2014). An aspirational framework for strategic human resource management. Academy of Management Annals, 8(1), 1-56.

Jackson, S. E., & Schuler, R. S. (2003). Managing human resources through strategic partnerships. South-Western Pub.

John, J. (2009). Study on the nature of impact of soft skills training programme on the soft skills development of management students. Pacific Business Review, 10(12), 19-27

Jackson and Schuler (2003) Managing Human Resources Through Strategic Partnerships

Johansson, I. (2004). Ontological Investigations: An Inquiry Into the Categories of Nature, Man and Soceity. De Gruyter

Kapoor, R. (2020). COVID-19 and the State of India's Labour Market. ICRIER policy series, 18(1), 1-7.

Kanfer, R., Wanberg, C. R., & Kantrowitz, T. M. (2001). Job search and employment: A personality–motivational analysis and meta-analytic review. Journal of Applied psychology, 86(5), 837.

Kehl, W., Jackson, M., & Fergnani, A. (2020). Natural language processing and futures studies. World Futures Review, 12(2), 181-197.

Kielkopf, C. F. (1977). Quantifiers in ontology. Studia Logica, 36, 301-307.

Knudsen, C., & Foss, N. J. (Eds.). (1996). Towards a competence theory of the firm. Routledge.

Ko, Y., & Seo, J. (2000). Automatic text categorization by unsupervised learning. In COLING 2000 Volume 1: The 18th International Conference on Computational Linguistics.

Konieczny, R., & Idczak, R. (2016). Mössbauer study of Fe-Re alloys prepared by mechanical alloying. Hyperfine Interactions, 237, 1-8.

Korde, V., & Mahender, C. N. (2012). Text classification and classifiers: A survey. International Journal of Artificial Intelligence & Applications, 3(2), 85.

Krishnaveni, J. (2013). A study on mapping of employees' competency. Indian Journal of Economics and Development, 1(3), 71-75.

Kumar, K., & Thakur, G. S. M. (2012). Advanced applications of neural networks and artificial intelligence: A review. International journal of information technology and computer science, 4(6), 57.

Lauriola, I., Lavelli, A., & Aiolli, F. (2022). An introduction to deep learning in natural language processing: Models, techniques, and tools. Neurocomputing, 470, 443-456.

Lawrence, D. (2006). Enhancing self-esteem in the classroom. Pine Forge Press.

Lee, S., Schmidt-Klau, D., & Verick, S. (2020). The labour market impacts of the COVID-19: A global perspective. The Indian Journal of Labour Economics, 63, 11-15.

Lee, T., Chun, J., Shim, J., & Lee, S. G. (2006). An ontology-based product recommender system for B2B marketplaces. International Journal of Electronic Commerce, 11(2), 125-155.

Ljp Van Der Maaten, L. (2014). Accelerating t-SNE using tree-based algorithms. The journal of machine learning research, 15(1), 3221-3245.

Ljp Van der Maaten, L., & Hinton, G. (2012). Visualizing non-metric similarities in multiple maps. Machine learning, 87, 33-55.

Ljp Van Der Maaten, L. (2009, April). Learning a parametric embedding by preserving local structure. In Artificial intelligence and statistics (pp. 384-391). PMLR..

Ljp Van Der Maaten & Hinton, G. (2008). Visualizing high-dimensional data using t-sne journal of machine learning research. J Mach Learn Res, 9, 26.

Lepold, A., Tanzer, N., & Jiménez, P. (2018). Expectations of bank employees on the influence of key performance indicators and the relationship with job satisfaction and work engagement. Social Sciences, 7(6), 99.

Liu, R., Feng, S., Shi, R., & Guo, W. (2014). Weighted graph clustering for community detection of large social networks. Procedia Computer Science, 31, 85-94.

Lo, A. (2017). Course Papers for Students of Business School Professional Degree Center 201 7 — 201 8. 0–6.

Lula, P., Oczkowska, R., Wiśniewska, S., & Wójcik, K. (2018). Ontology-based system for automatic analysis of job offers. Information Technology for Practice, 2018, 205-212.

Lula, P., Wiśniewska, S., & Wójcik, K. (2019). Analysis of the Demand for Competencies on the Polish Labour Market in the Context of Industry 4.0. In The 13th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings. Warszawa: Wydawnictwo CH Beck (pp. 124-131).

Lula, P., Oczkowska, R., Wiśniewska, S., & Wójcik, K. (2018). Ontology-based system for automatic analysis of job offers. Information Technology for Practice, 2018, 205-212.

Lula, P., Kovaleva, A., Oczkowska, R., Tyrańska, M., & Wiśniewska, S. (2019). Bipartite Competency Schemas on Polish Labor Market. Central European Business Review, 8(4), 1.

Lyman, E. W. (1914). Must Dogmatics Forego Ontology?. The American Journal of Theology, 18(3), 355-377.

M. Rajman,, & Besançon, R. (1998). Text mining: natural language techniques and text mining applications. In Data Mining and Reverse Engineering: Searching for semantics. IFIP TC2 WG2. 6 IFIP Seventh Conference on Database Semantics (DS-7) 7–10 October 1997, Leysin, Switzerland (pp. 50-64). Springer US.

Madhavi, T., & Mehrotra, R. (2019). Competency mapping of sales employees in pharmaceutical industry - A blue print for future. International Journal of Pharmaceutical Research, 11(1), 207–215.

Madhoushi, Z., Hamdan, A. R., & Zainudin, S. (2015, July). Sentiment analysis techniques in recent works. In 2015 science and information conference (SAI) (pp. 288-291). IEEE.

Madabushi, H. T., & Lee, M. (2016, December). High accuracy rule-based question classification using question syntax and semantics. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers (pp. 1220-1230).

Mali, M., & Atique, M. (2014). Applications of text classification using text mining. International Journal of Engineering Trends and Technology, 13(5), 209.

Malita, L. (2009). E-portfolios in an educational and ocupational context. Procedia-Social and Behavioral Sciences, 1(1), 2312-2316.

Majid, S., Liming, Z., Tong, S., & Raihana, S. (2012). Importance of soft skills for education and career success. International Journal for Cross-Disciplinary Subjects in Education, 2(2), 1037-1042.

Manasi, D. P. (2014). Improving effectiveness of retail sector through competency mapping of sales managers. Indian Journal of Applied Research, 4(10), 328-330.

Matthew, H., Holger, K., Alan, R., Robert, S., & Chris, W. (2004). A practical guide to building owl ontologies using the protege-owl plugin and co-ode tools.

Mawlawi, A., & Fawal, A. E. (2018). Talent Management in the Lebanese banking sector. Management, 8(3), 80–85.

McClelland, D. C., & Boyatzis, R. E. (1982). Leadership motive pattern and long-term success in management. Journal of Applied psychology, 67(6), 737.

McClelland, D. (1973). Testing for competence rather than intelligence. American Psychologist, 28, 1-14.

Messum, D., Wilkes, L., Peters, K., & Jackson, D. (2016). Content analysis of vacancy advertisements for employability skills: Challenges and opportunities for informing curriculum development. Journal of Teaching and Learning for Graduate Employability, 7(1), 72-86.

Miller, S. J. (2013, August). Introduction to ontology concepts and terminology. In DC-2013, Lisbon, Portugal.

Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into texts In Proceedings of EMNLP.

Mining, P. T. (2012). The Seven Practice Areas of Text Analytics. Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications, January, 29–41.

Ministry of Cabinet Affairs and the Future. (2014). UAE vision 2021.pdf (p. 25). www.vision2021.ae www.economy.ae. (n.d.).

Mirvis, P. H., Sales, A. L., & Hackett, E. J. (1991). The implementation and adoption of new technology in organizations: the impact on work, people, and culture. Human Resource Management, 30(1), 113-139.

Mitra, M., & Chaudhuri, B. B. (2000). Information retrieval from documents: A survey. Information retrieval, 2, 141-163.

Mohamad, M., Jamaludin, H., Zawawi, Z. A., & Hanafi, W. N. W. (2018). Determinants influencing employability skills: Undergraduate perception. Global Business and Management Research, 10(3), 568..

Mohammad, A. A. (2020). Understanding motivations, employability skills, employment aspiration, and training of hospitality management undergraduates. Tourism Review International, 24(4), 185-199.

Myers, M. B., Griffith, D. A., Daugherty, P. J., & Lusch, R. F. (2004). Maximizing the human capital equation in logistics: education, experience, and skills. Journal of business logistics, 25(1), 211-232.

Musen, M. A. (1992). Dimensions of knowledge sharing and reuse. Computers and biomedical research, 25(5), 435-467.

Nalawade, R. K., More, D. K., & Bhola, S. S. (2019). Employability skills required for functional areas of management. IUP Journal of Soft Skills, 13(1), 20-44.

Nelson, C. (2004). UAE national women at work in the private sector: Conditions and constraints. Center for Labour Market Research & Information.

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010, June). Automatic evaluation of topic coherence. In Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics (pp. 100-108).

Niharika, S., Latha, V. S., & Lavanya, D. R. (2012). A survey on text categorization. International Journal of Computer Trends and Technology, 3(1), 39-45.

Nguyen, D. D., Hagendorff, J., & Eshraghi, A. (2015). Which executive characteristics create value in banking? Evidence from appointment announcements. Corporate Governance: An International Review, 23(2), 112-128.

Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology.

Nusrat, M., & Sultana, N. (2019). Soft skills for sustainable employment of business graduates of Bangladesh. Higher Education, Skills and Work-Based Learning.

Olawale, Y. (2015). The employability skills provision within a construction project management degree programme.

Olorunnimbe, M. K., & Viktor, H. L. (2015). Tweets as a vote: Exploring political sentiments on twitter for opinion mining. In Foundations of Intelligent Systems: 22nd International Symposium, ISMIS 2015, Lyon, France, October 21-23, 2015, Proceedings 22 (pp. 180-185). Springer International Publishing.

Osoian, C., & Zaharie, M. (2014). Recruitment for competencies in public and private sectors. Transylvanian review of administrative sciences, 10(41), 129-145..

Oussii, A. A., & Klibi, M. F. (2017). Accounting students' perceptions of important business communication skills for career success: An exploratory study in the Tunisian context. Journal of Financial Reporting and Accounting.

Pater, R., Szkola, J., & Kozak, M. (2019). A method for measuring detailed demand for workers' competences. Economics, 13(1).

Papa, M. J. (1989). Communicator competence and employee performance with new technology: A case study. Southern Communication Journal, 55(1), 87-101.

Park, S., & Kim, W. (2007). Ontology mapping between heterogeneous product taxonomies in an electronic commerce environment. International Journal of Electronic Commerce, 12(2), 69-87.
Pavlinek, M., & Podgorelec, V. (2017). Text classification method based on self-training and LDA topic models. Expert Systems with Applications, 80, 83-93.

Payne, H. J. (2005). Reconceptualizing social skills in organizations: Exploring the relationship between communication competence, job performance, and supervisory roles. Journal of Leadership & Organizational Studies, 11(2), 63-77.

Petrović, S. (2007). Collocation extraction measures for text mining applications.

Pfeffer, J. (1994). Competitive advantage through people. Boston/Mass.

Pokrywczyński, D., & Malcolm, G. (2014). Towards a functional approach to modular ontologies using institutions. Studia Logica, 102, 117-143.

Quintana, C. D. D., Mora, J. G., Pérez, P. J., & Vila, L. E. (2016). Enhancing the development of competencies: The role of UBC. European Journal of Education, 51(1), 10-24.

Rajarathinam, V. K. (1970). Internet Banking Users's Competence and its Influence On Usage Satisfaction-A View from India. The Journal of Internet Banking and Commerce, 18(3), 1-13.

Ra. Baeza-Yates, B. R. N. R. R. baeza-yates and b. ribeiro-neto: Modern information retrieval, addison wesley (1999). *vol*, *17*, 110-110.

Rani, N., & Singla, J. (2015). Auditing human resource functions & competencies: an empirical study. Indian Journal of Industrial Relations, 109-120.

Rao, T., & Srivastava, S. (2014). Twitter sentiment analysis: How to hedge your bets in the stock markets. State of the art applications of social network analysis, 227-247.

Rashmi, K., & Singh, R. (2020). Building competitive advantage through engagement of employees: A conceptual model. The International Journal of Interdisciplinary Organizational Studies, 15(1), 1.

Revolution, F. I. (2020). Thrive during Fourth Upskilling Secondary School Learners. 6(June), 3–12.

Roberts, G. (1997). Recruitment and selection. CIPD publishing.

Robertson, S. E., & Jones, K. S. (1976). Relevance weighting of search terms. Journal of the American Society for Information science, 27(3), 129-146.

Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2012). The author-topic model for authors and documents. arXiv preprint arXiv:1207.4169.

Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. Text mining: applications and theory, 1-20..

Razavi, R. (2020). Personality segmentation of users through mining their mobile usage patterns. International Journal of Human-Computer Studies, 143, 102470.

Yuvaraj, R. (2011). Competency mapping. International journal of scientific & engineering research, 2(8), 1-7.

Sabban, R. (2002). United Arab Emirates: Migrant Women in the United Arab Emirates. The Case of Female Domestic Workers. Gender Promotion Program.

Salton, G. (1963). Some hierarchical models for automatic document retrieval. American Documentation, 14(3), 213-222.

Salton, G., Wong, A., & Yang, C. S. (1975). Vector Space Model for Automatic Indexing. Information Retrieval and Language Processing. Communications of the ACM, 18(11), 613–620.

Salman, M., Ganie, S. A., & Saleem, I. (2020). Employee competencies as predictors of organizational performance: a study of public and private sector banks. Management and Labour Studies, 45(4), 416-432.

Schiliro, D. (2013). Diversification and development of the UAE's economy.

Schütze, H., Manning, C. D., & Raghavan, P. (2008). Introduction to information retrieval (Vol. 39, pp. 234-265). Cambridge: Cambridge University Press.

Shayah, M. H. (2015). Economic diversification by boosting non-oil exports (case of UAE). J. Eco. Bus. Manage.(JOEBM), 3(7), 735-738.

Sharma, A., Adhikary, A., & Borah, S. B. (2020). Covid-19' s impact on supply chain decisions: Strategic insights from NASDAQ 100 firms using Twitter data. Journal of business research, 117, 443-449.

Sharma, A., Adhikary, A., & Borah, S. B. (2020). Covid-19' s impact on supply chain decisions: Strategic insights from NASDAQ 100 firms using Twitter data. Journal of business research, 117, 443-449.

Shafie, L. A., & Nayan, S. (2010). Employability awareness among Malaysian undergraduates. International journal of business and management, 5(8), 119.

Shen, C. W., Chen, M., & Wang, C. C. (2019). Analyzing the trend of O2O commerce by bilingual text mining on social media. Computers in Human Behavior, 101, 474-483.

Sievert, C., Shirley, K., & Davis, L. A method for visualizing and interpreting topics. In Proceedings of Workshop on Interactive Language Learning, Visualization, and Interfaces, Association for Computational Linguistics (pp. 63-70).

Singh, A. (2012). Job satisfaction among the expatriates in the UAE. International Journal of Business and Social Research, 2(5), 234-249.

Singh, S. (2019). Building Employability Skills in English as a Second Language (ESL) Classroom in India. English Teacher, 48(2).

Singhal, V., & Saini, R. (2020). Education And Its Contribution In Enhancing Employability Skills: Study On The Efficacy Of Management Education. PalArch's Journal of Archaeology of Egypt/Egyptology, 17(9), 1074-1090.

Sibarani, E. M., Scerri, S., Morales, C., Auer, S., & Collarana, D. (2017, September). Ontology-guided job market demand analysis: a cross-sectional study for the data science field. In Proceedings of the 13th International Conference on Semantic Systems (pp. 25-32).

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. Journal of documentation, 28(1), 11-21.

Spencer, L. M., & Spencer, P. S. M. (2008). Competence at Work models for superior performance. John Wiley & Sons.

Jones, K. S., Walker, S., & Robertson, S. E. (2000). A probabilistic model of information retrieval: development and comparative experiments: Part 2. Information processing & management, 36(6), 809-840.

Steede, G. M., Meyers, C., Li, N., Irlbeck, E., & Gearhart, S. (2018). A sentiment and content analysis of Twitter content regarding the use of antibiotics in livestock. Journal of Applied Communications, 102(4), 1B-1B.

Stevenson, D. H., & Starkweather, J. A. (2010). PM critical competency index: IT execs prefer soft skills. International journal of project management, 28(7), 663-671.

Suleman, F. (2018). The employability skills of higher education graduates: insights into conceptual frameworks and methodological options. Higher Education, 76, 263-278.

Suarta I. M., Suwintana I. K., Sudhana I. F. P., & Hariyanti N. K. D. (2018) Employability skills for entry level workers: a content analysis of job advertisements in Indonesia, Journal of Technical Education and Training, 10 (2), 49-61.

Sugden, R. (2016). Ontology, methodological individualism, and the foundations of the social sciences. Journal of Economic Literature, 54(4), 1377-1389.

Shweta Chaudhary and Seema Singh (2016) “A study of competency mapping of employees in banking sector” international Journal of HRM vol r Issue 3 11-20

Tan, H. H. (2009). Firm–employee relationship strength—Competitive advantage through people revisited: A commentary essay. Journal of Business Research, 62(11), 1108-1109.

Tsoumas, B. and Gritzalis, D. “Towards an Ontology-based Security Management”, in the Proceedings of the 20th International Conference on Advanced Information Networking and Application, 2006

UNDP. (2015). Human Development Report 2015: Work for Human Development (Technical Notes).

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11).

Van Rijsbergen, C. J. (1986). A non-classical logic for information retrieval. The computer journal, 29(6), 481-485.

van Klink, M. R. D., & Boon, J. (2003). Competencies: The triumph of a fuzzy concept. International Journal of Human Resources Development and Management, 3(2), 125-137.

Varghese, B., & Govilkar, S. (2015). A survey on various word spotting techniques for content-based document image retrieval. International Journal of Computer Science and Information Technologies, 6, 2682-2686.

Vikram C. & Srivastava, S. (2013). Competency mapping for HR professionals in IT industry. The international journal of management, 2(3), 1-6.

Vis, B. N. Cities Made of Boundaries. biography, 90(93), 161.

Velicanu, A., Lungu, I., Diaconita, V., & Nisioiu, C. (2013). The 9th International Scientific Conference eLearning and software for Education.

Wu, T., Hao, F., & Kim, M. (2021). Typical opinions mining based on Douban film comments in animated movies. Entertainment Computing, 36, 100391.

Wang, C., Zhu, H., Wang, P., Zhu, C., Zhang, X., Chen, E., & Xiong, H. (2021). Personalized and explainable employee training course recommendations: A bayesian variational approach. ACM Transactions on Information Systems (TOIS), 40(4), 1-32.

Weiss, S. M., Indurkhya, N., & Zhang, T. (2015). Fundamentals of predictive text mining. Springer.

McIlvaine, A.R. (1998). 'World Premiere', Human Resource executive, 19, October, pp. 18–20.

Wheelahan, L. (2007). How competency-based training locks the working class out of powerful knowledge: A modified Bernsteinian analysis. British Journal of Sociology of Education, 28(5), 637-651.

Wang, Y. F., & Tsai, C. T. (2014). Employability of hospitality graduates: Student and industry perspectives. Journal of Hospitality & Tourism Education, 26(3), 125-135.

Wesslen, R. (2018). Computer-assisted text analysis for social science: Topic models and beyond. arXiv preprint arXiv:1803.11045.

Wieczorek-Szymańska, A. (2015). Employees' Competencies Management in Bank Sector. Reports on Economics and Finance, 1(1), 105-113.

Whiddett, S., & Hollyforde, S. (2007). Competencies. London: CIPD.

Whiddett, S., & Hollyforde, S. (2003). A practical guide to competencies: how to enhance individual and organisational performance. CIPD Publishing.

Woodruffe, C. (1991). Competent by any other name. Personnel Management, 23(9), 30-33.

Wright, P. M., McMahan, G. C., & McWilliams, A. (1994). Human resources and sustained competitive advantage: a resource-based perspective. International journal of human resource management, 5(2), 301-326.

Yamarone, R. (2012). The trader's guide to key economic indicators. John Wiley & Sons.