# Applying Workforce Intelligence Clustering on Egyptian Governorates

Nevine M. Labib
Computer and Information Systems Department,
Sadat Academy for Management Sciences,
Cairo, Egypt
nevmakram@gmail.com

Ahmed E. Sabry
Computer and Information Systems Department,
Sadat Academy for Management Sciences,
Cairo, Egypt
aesmat@sadatacademy.edu.eg

**Full/Regular Research Paper**
**CSCI-ISAI**

***Abstract*—Workforce Intelligence incorporates various tools that capture and record details about the organizations' workforce and relevant activities, such as employees' skills, experience, education, job description and other demographic data. In addition, it correlates the enterprises' operational systems with workforce activities to increase the efficiency of the analysis process and rationalize decision-making. It has several usages such as workforce planning and optimization, transitions, analytics and productivity.**

**Since Egypt is currently focusing on the human capital because of its vital role in achieving Egypt's 2030 vision, this study analyzes national workforce data, classified according to governorates, and applies unsupervised clustering technique to portray employees' profile. The research conducts profiling of the twenty-seven Egyptian governorates using multi-dimensional mining with fifty-four dimensions. It addresses workforce distribution of employment, unemployment, industries and educational levels data. The resulted clustering attributes will help in the workforce planning and optimization in Egypt.**

***Keywords—Workforce Intelligence, Clustering, Business Intelligence, Data Mining, Employee Profiling***

## I. INTRODUCTION

### A. Background

Egypt vision 2030 is about "creating a modern, open, democratic, productive, and happy society**" [1].** It aims at ***"An efficient and effective government administrative body, characterized by professionalism, transparency, justice and responsiveness, offers quality services, and can be held accountable, able to increase citizen satisfaction, and strongly contributes in the achievement of Egypt's development goals and in improving the status of the Egyptian people."*** As for the main goals of the Sustainable Development Strategy (SDS) to achieve this vision, they include 'Human Capital' and 'Competitive Market' [2]. Accordingly, one of the SDS pillars is *Education and Training,* targeting administrative management and reform.

A main challenge in achieving this vision is human development since human resources play a major role in this change. Therefore, it is crucial to develop a framework that integrates available workforce data collected from different governmental entities, and analyze these data using several intelligent techniques in order to extract indicators to be used for workforce planning and development.

### B. Problem Definition

Although many government entities have considerable data collected from various forms of transactions, the following problems are encountered:

a. There is no standardization / unification of data models.
b. Not all required data are available.
c. There is no integration between the available data.

On the other hand, these data should be analyzed thoroughly in order to provide useful indicators that rationalize decisions towards strategic directions. This analysis will make use of several techniques such as Business Intelligence, and Data Mining, hence providing Workforce Intelligence support for organizational development and reform. This paper is an initiative towards proposing a framework for such integrated analysis.

### C. Originality

The originality of the study stems from integrating data collected from different sources and having different formats, and multi-dimensional clustering for these data using demographic and geographic dimensions then analyzing these clusters in order to extract meaningful indicators that can be used for workforce planning and optimization Egypt.

### D. Objectives

The study aims at making use of Workforce Intelligence analytics to study employees profiling in Egypt.

1.4.1 Main Objective:

The main objective of the study is to propose a national / governmental unified framework that incorporates data from various sources so as to allow simple integration in addition to providing relevant indicators for effective planning of future workforce requirements.

978-1-5090-5510-4/16 $31.00 © 2016 IEEE
DOI 10.1109/CSCI.2016.105

*1) Detailed Objectives:*

The study has also other detailed objectives from strategic management perspective. They are as follows:

- Identifying the need for new business directions or domains.
- Determining which business domains or specializations can be reassigned or eliminated.
- Determining required education levels and trends.
- Identifying unemployment distribution over governorates.
- Identifying employment distribution over governorates.

### E. Importance

The importance of the study lies in providing an integrated view of employees' profiles in different governorates, taking into consideration several dimensions, such as educational level and industry. This profiling will assist policy makers in taking suitable decisions related to strategic workforce planning and optimization, which will eventually contribute to administrative reform in Egypt, since it is the core of Egypt's vision 2030.

## II. Workforce Intelligence Concept and AnalyticsUse

### A. Workforce Intelligence

Most organizations have large quantities of employee historical data archived within their operational systems. These data can be used to extract meaningful patterns that help in the management of human resources in the workplace, hence increasing their efficiency, by means of workforce intelligence solutions. These solutions provide a comprehensive set of functionality that captures and records detailed profiles of the organization employees, including their qualifications, skills, experience, and competencies in addition to their daily activities**[3**]. They also make use of other sources of data such as Enterprise Resource Planning, Supply Chain Management and Customer Relationship Management. These integrated data are then analyzed using different techniques in order to extract new knowledge and make it available for decision makers and management who rely on it when planning for their workforce. These techniques include business intelligence, data mining and predictive analysis.

### B. Business Intelligence and Analytics

Business Intelligence (BI) is a broad term that indicates using data to uncover new patterns that support the decision-making process. BI functions include the collection, analysis, visualization, and reporting of data. Hence, it involves both technical and managerial aspects, since the management has to direct the analyzed data towards more rational decisions [4].

One of the recent trends in this domain is 'self-service BI', which are tools used by business users in order to generate custom reports and visualizations and disseminate with other stakeholders, in order to support them in selection of the course of actions to take, without the need of IT professionals[4].

While BI is generally about data collection, aggregation, visualization and reporting, with the goal of supporting decisions or business analytics (BA) is about asking questions whose answers lead to the decision-making process. These questions are related to breaking down the available data, and making relevant comparisons. Therefore, the questions can be about the past or the future [5].

On the other hand, data mining (DM) is about discovering new and hidden patterns, that were previously unknown and that are important to the decision makers. It makes use of several disciplines such as statistics and artificial intelligence, and data science [6].

Finally, the previously discussed tools are considered the pillars of the workforce analytics, aiming at knowledge and pattern discovery in addition to providing meaningful interpretation used for prediction and planning.

### C. Role of Workforce Analytics

Workforce analytics have a major role in helping managing the human resources of the organization. They can help in the decisions concerning recruitment, development and retention of employees. They can also help in workforce planning by identifying the need for new departments and positions, identifying required education trends, identifying factors that influence employee job satisfaction, predicting current and future technological needs, and optimizing the enterprise's organizational structure, hence, helping in the developing of the organization [7].

## III. Data and Framework

### A. Sources of Data

Data were gathered from various sources including:

- The Egyptian Central Agency for Public Mobilization and Statistics (CAPMAS),
- The Egyptian Ministry of Manpower (MOP), and
- The Egyptian Ministry of Planning, Monitoring and Administrative Reform (MOPMAR).

These data were cleaned, aggregated and classified according to governorates, in order to be used in the study.

Fig. 1 illustrates the Egyptian governorates administrative boundaries map. It will be used later in the study to visualize governorates' codes and resulting clusters.

### B. Data Dimensions

The data are related to 2015. They include the following dimensions:

- Employment (number of employed persons above 15 years old) classified based on gender, industry, and governorate. It includes 247,789 records.
- Population distribution of 2015 according to the Central Agency for Public Mobilization and Statistics (published in March 2016), includes 89,883,377 records and covers 27 Governorates (1053953 km2). Later, within 2016 the number of governorates changed as they increased to 30.
- Distribution of group complaints according to governorates in 2015. It includes 1,173 records.

- Distribution of holders of University degree job seekers according to governorates, educational level and gender in 2015. It includes 1,694,717 records.
- Distribution of non-holders of University degree job seekers distributed according to governorates and gender in 2015. It includes 979,663 records.
- Requests of individual disputes settlement distributed according to governorates and adopted from published report in 2015. It includes 45,550 records.
- Samples Grand Total: 93,906,332 records.

Following are the tables listing the main dimensions categories and their corresponding codes used in the analysis phase. The main categories of industries used in the analysis are listed in Table I.

TABLE I. INDUSTRY CATEGORIES USED IN DATA ANALYSIS

| Code | Industry Description |
|---|---|
| A | A Agriculture, Hunting , Forestry and Cutting of wood trees |
| B | B Mining and quarrying |
| C | C Transformational Manufacturing |
| D | D Electric, gas, steam, and air condition supplies |
| E | E Water supply and Sanitation |
| F | F Construction and Building |
| G | G Whole and retail sale vehicles and motorcycles repairing |
| H | H Transportation and Storage |
| I | I Food and Residence services |
| J | J Information and Telecommunications |
| K | K Insurance and Financial Intermediation |
| L | L Real estate and Renting |
| M | M Specialized technical and scientific |
| N | N Administrative Activities & Support Services |
| O | O Public Administration ,defense, social solidarity |
| P | P Education |
| Q | Q Health and Social Work |
| R | R Entertainment &Creation &Arts Activities |
| S | S Other Services Activities |
| T | T Services of home service for private households |
| U | U International and Regional Agencies &Organizations |
| V | V. Unclassified Activities |

As for the data available for raised complaints during 2015 contains total of 1,173 recorded submitted complaints, they were categorized and used in the study as listed in Table II.

TABLE II. EMPLOYEES' COMPLAINTS CATEGORIES

| Code | Complaint Category | Code | Complaint Category |
|---|---|---|---|
| 1 | Wages | 13 | Investigated Cases |
| 2 | Group Complaints about Wages | 14 | Other Entities |
| 3 | Transfer to Labor Court | 15 | No Action |
| 4 | Previous Complaints | 16 | Social Services |
| 5 | Insurance Notice Group Complaints | 17 | Working Hours |
| 6 | Social Insurance Notice | 18 | Firing |
| 7 | Inspection Notice | 19 | Under Investigations |
| 8 | Inspection Notice | 20 | Firing |
| | Settlements based on investigations | | |
| 9 | Friendly Settlements | 21 | Transferred Requests |
| 10 | Friendly Settlements based on investigations | 22 | Requests |
| 11 | Total Submitted and Carryover | 23 | Other requests |
| 12 | Submitted and Carryover Group Complaints | 24 | Annual Grants |

### C. *Governorates Clusters Used in Data Analysis and Visualization*

Following Fig. 1. Represents the Egyptian governorates map and their relevant codes. These codes will be used for the analysis purposes and referred to in the clusters labeling.

TABLE III. LIST OF EGYPTIAN GOVERNORATE NAMES WITH CORRESPONDING CODES

| Code | English Name | Code | English Name |
|---|---|---|---|
| 1 | Aswan | 15 | Qalyubia |
| 2 | Asyut | 16 | Monufia |
| 3 | Alexandria | 17 | Minya |
| 4 | Ismailia | 18 | New Valley |
| 5 | Luxor | 19 | BeniSuef |
| 6 | Red Sea | 20 | Port Said |
| 7 | Beheira | 21 | South Sinai |
| 8 | Giza | 22 | Damietta |
| 9 | Dakahlia | 23 | Sohag |
| 10 | Suez | 24 | North Sinai |
| 11 | Sharqia | 25 | Qena |
| 12 | Gharbia | 26 | Kafr el-Sheikh |
| 13 | Faiyum | 27 | Matruh |
| 14 | Cairo | | |

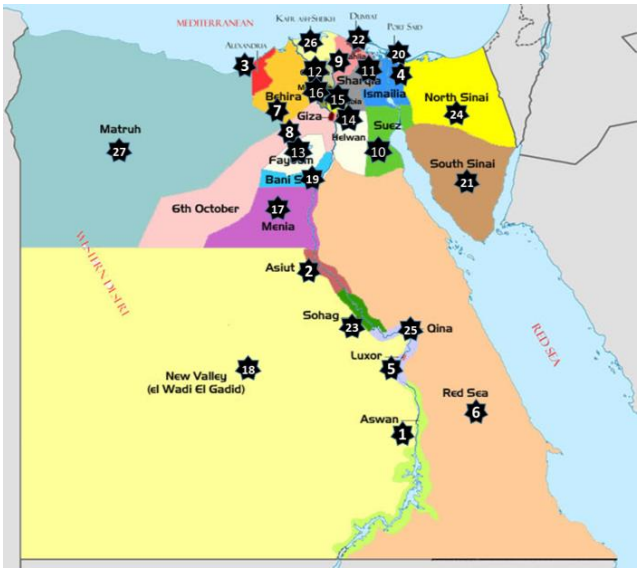


Fig. 1. Egyptian Governorates and their Corresponding Codes

### D. *Tool Used for Analysis*

The tool used for data analysis in the study is RapidMiner. It is a complete business analytics workbench with a strong focus on data mining, text mining, and predictive analytics [6]. It uses a wide variety of both descriptive and predictive techniques so as to provide insights to help with decision making and offers full reporting and dashboarding capabilities and, therefore, a complete business intelligence solution in combination with predictive analytics. It also provides data preprocessing and visualization, predictive analytics and statistical modeling, evaluation, and deployment. The implemented mining model of Rapid Miner tool is illustrated in Fig 2.

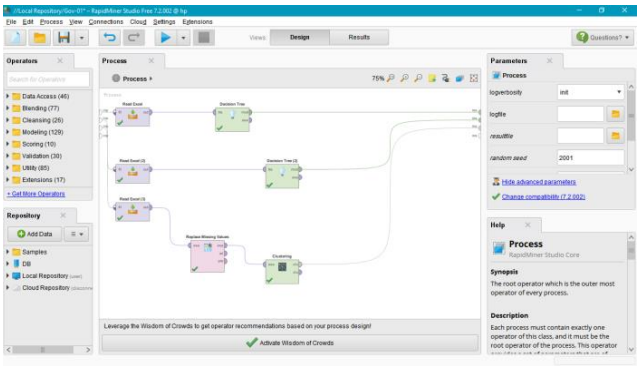

Fig. 2. Designed Mining Model on RapidMiner Tool

## IV. CLUSTER PROFILING OF THE WORKFORCE IN EGYPT

### A. *Detailed Description of the Clusters*

This section discusses the visualized clusters' results on Egyptian Governorates map (Administrative Classification). The aggregated data are analyzed into five clusters according to governorates. Following are the clusters along with the details of the dimensions used listed in Table IV.

TABLE IV. RESULTED CLUSTERS MAPPED TO CODES OF EGYPTIAN GOVERNORATES

| Governorate Codes | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
|---|---|---|---|---|---|
| Governorate 1 | 2 | 1 | 13 | 8 | 7 |
| Governorate 2 | 3 | 4 | 19 | 14 | 9 |
| Governorate 3 | 12 | 5 | 25 | - | 11 |
| Governorate 4 | 15 | 6 | 26 | - | - |
| Governorate 5 | 16 | 10 | - | - | - |
| Governorate 6 | 17 | 18 | - | - | - |
| Governorate 7 | 23 | 20 | - | - | - |
| Governorate 8 | - | 21 | - | - | - |
| Governorate 9 | - | 22 | - | - | - |
| Governorate 10 | - | 24 | - | - | - |
| Governorate 11 | - | 27 | - | - | - |
| No of Mapped Governorates | 7 | 11 | 4 | 2 | 3 |

The coded industries mapped to the resulted clusters are shown in Fig 4.

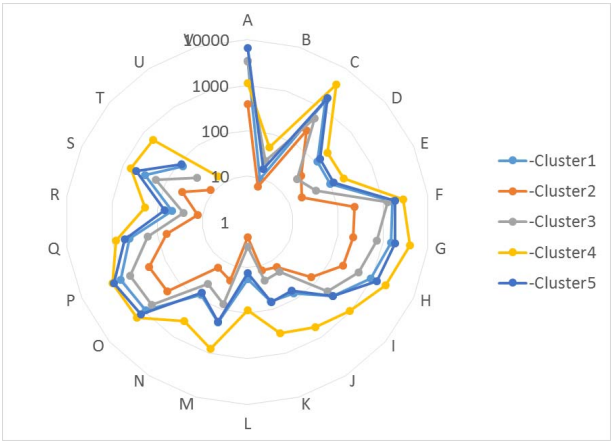


Fig. 3. Industries mapped to the resulted clusters

Table V and Fig 5 illustrate the different age and education levels used in unemployment data handling within different governorates.

TABLE V. AGE AND EDUCATION LEVELS USED IN UNEMPLOYMENT DIMENSION

| Level Code | Level Description |
|---|---|
| Level 1 | Uneducated, Young |
| Level 2 | Uneducated, Non-technical |
| Level 3 | Uneducated, Technical |
| Level 4 | Educated, Primary |
| Level 5 | Educated, Preparatory |
| Level 6 | Educated, High school |
| Level 7 | Educated, Graduate |
| Level 8 | Uneducated, Overaged |

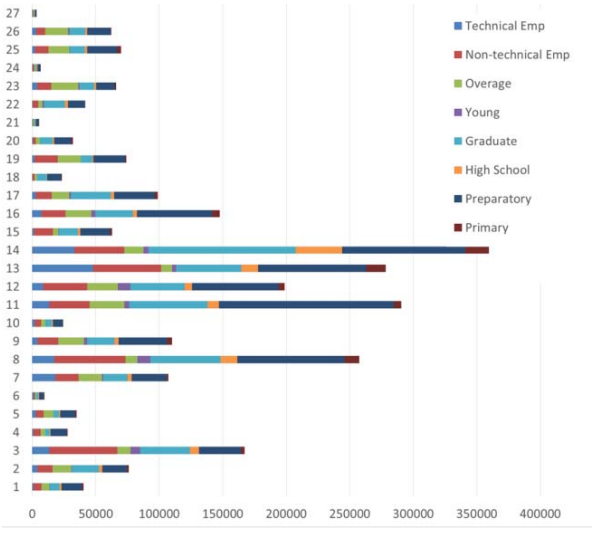


Fig. 4. Age and Education Levels of Unemployed per Governorate

## B. *Main Findings of the Study*

The resulted clusters mapped according to age and educations levels are shown in Fig 6.

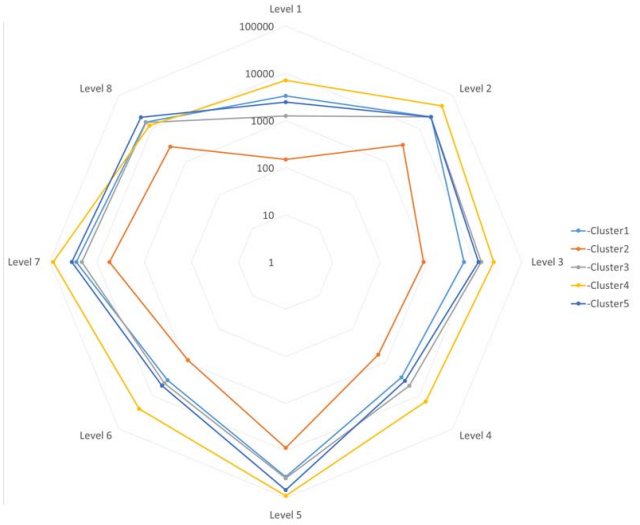


Fig. 5. Requests for Employment by Age and Education Levels Mapped to Resulted Clusters

The complaints' types mapped to the resulted clusters are illustrated in Fig 7.

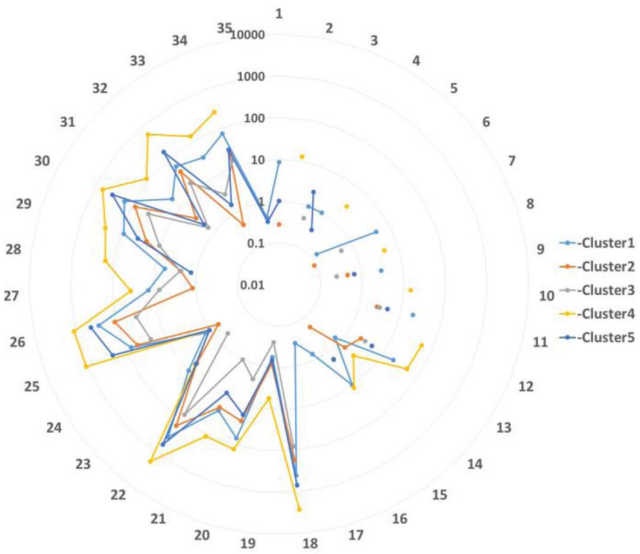


Fig. 6. Employees' Complaints Categories Mapped to the Resulted Clusters

The five clusters visualized on the Egyptian governorates map are shown in Fig 8. The map is followed by a detailed description of the characteristics of each cluster, such as main industries, complaint types and results, and unemployment age and educational levels.

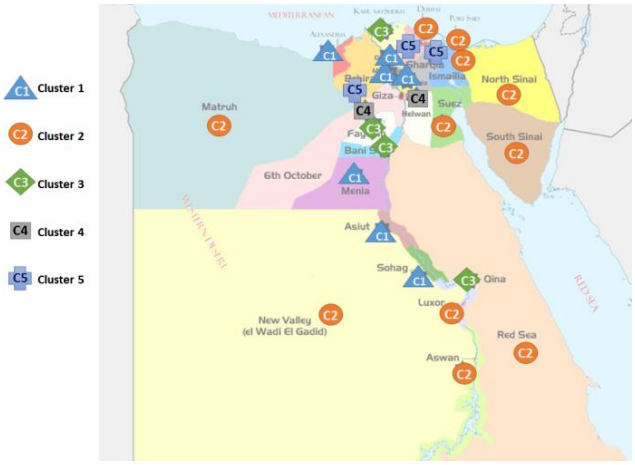


Fig. 7. Clusters Visualization on the Governorates Map

TABLE VI. CLUSTERS RESULTS DESCRIPTION

| Cluster | Main Complaint Categories | Complaints Results | Industries abbreviation (in order) | Unemployment Age and Education Levels | Other Factors |
|---|---|---|---|---|---|
| **Cluster 1** | • Social Services<br>• Working Hours<br>• Wages<br>• Work Process<br>• Others | • Inspection Notice<br>• No action | • J, L, and A | • Kids<br>• Non-technical Employees<br>• Elderly People | • Large Population |
| **Cluster 2** | • Nothing specific | • Nothing specific | • Nothing specific | • Below average education | • Large Area |
| **Cluster 3** | • Nothing specific | • Nothing specific | • B , and A | • Below average education<br>• Technical Employees<br>• Elderly people | • Nothing specific |
| **Cluster 4** | • Group complaints<br>• Wages<br>• Firing<br>• Individual complaints<br>• Requests | • Settlement<br>• Inspection Notice<br>• Firing<br>• Other requests<br>• Settlement of Firing complaints<br>• Under investigation<br>• Settlement for requests<br>• No action<br>• Transfer to labor court<br>• Firing-No action<br>• No action | • U , J , T , L , K , N , M , R , I , C , G , B , Q , S , H , D , E , F , and O | • High school<br>• Kids<br>• University Graduates<br>• Technical Employees<br>• Non-technical Employees | • Large Population<br>• Large Area |
| **Cluster 5** | • Work process<br>• Firing | • Settlement of Firing complaints<br>• Firing-transferred to labor court | • K , R , I , C , G , A , Q , S , H , F , and O | • University Graduates<br>• Elderly people | • Large Population |

## V. SUMMARY AND CONCLUSION

### A. Implications of Main Findings

Based on the previous detailed labeling, interpretation and characteristics of each cluster, there are some findings shall be considered as follows:

***Cluster 1***: Results indicate that these governorates shall focus on providing solutions to decrease unemployment as training and direct education to empower more people to join the main industries within this cluster. In addition, focus on IT and smart cities may be developed in order to provide new job opportunities.

***Cluster 2:*** Governorates of this cluster shall focus on different education types and levels to improve the quality of human capital.

***Cluster 3:*** Results indicate that these governorates shall work on training programs to enhance the available technical skills of employees so as they contribute to the existing industries.

***Cluster 4:*** These governorates shall focus on administrative development and solving managerial issues, as it shows higher unresolved complaints, which may cause increase in unemployment rate within different ages and educational levels.

***Cluster 5:*** Governorates of this cluster shall focus on human capital development programs to enable resources to join different current and new suitable industries.

To sum up, based on the extracted results, various decisions may be considered, such as decisions concerning recruitment, development and retention of employees. In addition to focus on a specific set of industries, that suits each cluster governorates.

### B. Importance of Governorates Clustering

These clusters help to recognize the governorates' workforce so as to be able to develop their skills, create specialized workforce, train people for employment, engage community-based employment intermediaries, ones to fill gaps, and finally support regional skills.

All of the previous decisions should act as a support for administrative reform of the public sector at the local level; which will eventually support decentralization.

On the other hand, it may facilitate curtailment of the public sector, i.e., moving people from the public to the private sector, which requires identifying the absorptive capacity of the different sub-sectors within the private sector.

## VI. RECOMMENDATIONS AND FUTURE DIRECTIONS

### A. Recommendations based on the study

Following are some recommendations based on the study results.

- Workforce Planning: Plan for future needs based on both workforce demographic and geographic data.
- Workforce Cost Planning: Provide workforce data in order to support accurate planning, and continuous monitoring of actual performance relative to plan.
- Workforce Benchmarking: Measure standard workforce processes and compare the measurements with external benchmarks and internal operating thresholds.
- Workforce Process Analytics and Measurement: Measure and analyze typical core HR processes, such as payroll, time management, and benefits in addition to organizational structures, relationships, and attributes of jobs and positions.
- Talent Management: Analyze employee skills and qualifications and measure the effectiveness of training programs.
- Strategic Alignment: Ensure that all business activities are in line with the strategic goals of the organization, whether public or private. These strategies should contribute to the main strategy of the country.
- The extracted results may also help identifying the need for new departments and positions, as well as identifying required education trends. They can also identify factors that influence employee job satisfaction, if provided with more details concerning employee complaints.

### B. Fuure Directions

In order to achieve Egypt's vision 2030, governmental organizations need to create a position of data scientists, and analytical business intelligence professionals, who are capable of developing and utilizing novel algorithms and quantitative models.

As for future work, it is focused mainly on making use of Cloud Computing and big data analytics since the data required for such a model is huge and needs special handling when it comes to analysis, management and storage.

## *References*


[1] N. El-Megharbel. "Sustainable Development Strategy: Egypt's Vision 2030", ,New York, May 27, 2015.

[2] www.mof.gov.eg/MOFGallerySource/English/SME/Research_studies /10.pdf [Accessed October 2016]

[3] M. Butler. "Workforce Intelligence, Turning Workforce Data into Workforce Efficiency", Martin Butler Research 201, March 2014.

[4] "The Small and Medium Enterprises Policy Development Project (SMEPol)," Ministry of Foreign Trade, Cairo, Egypt, 2003.

[5] B. Kapoor. "Business Intelligence and Its Use for Human Resource Management", The Journal of Human Resource and Adult Learning Vol. 6, Num. 2, pp 21-30, December 2010.

[6] N. Agarwal. "Cluster Profiling of the Employees based on Skill Typology," 2015. [Online]. Internet: http://shodhganga.inflibnet.ac.in/ bitstream/10603/49332/18/18_annexure6.pdf. [Accessed October 2016].

[7] H. Chen, R.H.L. Chiang, and V.C. Storey. "Business Intelligence and Analytics: From Big Data to Big Impact", MIS Quarterly, Vol. 36, No. 4, pp. 1165-1188, December 2012.