Title: Employment Prediction using Machine Learning and Natural Text Processing Based on Current Job Market in OMAN
Proposal ID: BFP/RGP/ICT/18/091
Type of project application: Research Grant Program

Current Status: Approved For Funding

Sector Name
Information Technology and Communication

Principal Investigator
Gupta, Dr Sanjai (sanjay.gupta@nct.edu.om)

Co-Prinicipal Investigator
Faisal, Dr. Mohammed (mohammed.faisal@nct.edu.om)

Co-Investigators

Other Team Members
Team members not specified

Technology Readiness Level (TRL)
4 – 6: Technology Development/ Demonstration (public & private collaboration) e.g. prototype testing, advanced-stage research

Priority Area
Other

Executive Summary
The research proposal is lighting the unemployment in the Oman. The major problem for a recruiter is to filter out the candidates and choose the right one based on the required skill set. Finding the eligible candidates is never an issue but the abilities of the eligible candidate's is questionable, in sequence of that the proposal will move around to reach the final goal to facilitate the Omani job providers and academic institutions to furnish the skilled candidates who will be having eligibility and abilities to pursue their work with the required skill set.

In continuation, research proposal will focus on three major entities; academic institution of Oman, industries and the Omani scholars. Firstly, Big Data Analytics will be executed on the historical data collected from the various academic institutions and industries of Oman, to find the available candidates, their skill sets to know either it is matching with the current need, and required skill set of the employers.

Secondly, the data analytics will be done on the primary dataset collected from social media posts and various job portals in order to find the current job trend and level of satisfaction from the employer perspectives. The research deliverable will also produce the reason of unemployment and recommend the solution to minimize the raised issue.
Finally, we will construct machine-learning model for predictive analysis where training and test data set will be bifurcated from the primary dataset. Based on the machine learning deliverables future job availability, required skill sets will be forecasted which could be used for the deployment purpose.

Introduction and Statement of the Problem / Project

Unemployment is one of the major problems of any country. Every country needs to find the way to solve this problem. As per trading economics magazine “Unemployment Rate in Oman expanded to 16 percent in 2017 from 15.80 percent in 2016”. Joblessness Rate in Oman arrived at the midpoint of 18.15 percent from 1991 until 2017. [1]
Every year Oman is producing many graduates, which are lacking in required skills and knowledge to get suitable job. Employers are looking for skills and knowledge among the graduates so that employers need not to provide them additional training to acquire required skills.
In our research, we will study the condition of unemployment among the Omani graduates from the college of Technologies, Universities and different academic institutions of Oman. Our research objective is to collect and study the primary and secondary data about current job market, required skill sets, current curriculum of academic institutions of Oman in order to find out the reasons behind unemployment among Omani graduates and also find the possible solutions to minimize the unemployment among Omani graduates.
The main entities of our research will be Industry of Oman, current students, recent Omani graduates, job portals in Oman, social media posts, College of technologies, Universities and other academic institutions of Oman.
In this research, we will use Big data analytics tools for the collection of primary data from various sources like social media, job portals of Oman and from other concern departments. Using these data, we will use different analytics to find the information about job availability, required skill sets and g gaps in the curriculum of academic institutions of Oman.
In our research, we will use machine-learning analytics for the employment prediction, trending skill sets and availability of Omani graduates as per the job availability.
In this research project, we will analyze the big volume of social networking data and weblog data from different job portals using Apache Spark, Apache Hadoop ecosystem and rapid miner. Using this big volume of primary and secondary data, we will do the different analysis, which could be Diagnostic, Prescriptive, and Predictive .These different analysis will be used to get the useful information to minimize the unemployment among the Omani graduates.

Literature Review and Analysis of Related Work

1. Murali K. Pusala, Mohsen Amini Salehi (n.d.) (2016) has investigated best in class innovations in programming, registering, and capacity of the enormous information investigation scene. They explore on various sorts of examination that can be performed on monstrous information. For that, these researchers initially give a point-by-point scientific categorization on various expository sorts

alongside instances of each sort. Next, we feature innovation patterns of monstrous information investigation that are accessible for partnerships, government organizations, and specialists.[2]
2. Annina Simon , Mahima (n.d.) (2016) have discussed about the deep learning in area of Machine learning which can be used in applications of Big data analytics and artificial intelligence. In this paper author has discussed about two types of Machine learning tasks: Supervised and Unsupervised Machine learning. Deep learning is just a little venture towards building machines which have human-like knowledge. Further progressions must be made so as to accomplish our definitive objective. Associations like Google, Facebook, Microsoft and Baidu (a Chinese web index) are getting tied up with this innovation and investigating different roads accessible.[3]
3. Patrick Hall, Navdeep Gill (n. d.) (2018) has discussed about the practical techniques for Interpreting Machine Learning models using Python, apache ecosystem. In this paper author has given the description of the following models: Monotonically constrained GBMs, partial dependence, and ICE and Decision tree surrogate models. These models can be used to enhance understanding of the complex data and also enhance the accountability of its predictions.[4]
4. Abhishek Kori(2017) has discussed about data mining using Rapid miner. In this paper has explained how we can extract predictive information from the large database In this paper author has discussed various algorithms like Naïve Bayes classification using rapid miner tool. In his paper author has carried out his experiment using KNN which gives more accurate result compare to Naïve Bayes.[5]
5. Thirunavukkarasu K,Dr.Manoj Wadhawa(2016) have compared different data mining algorithms such as :Learning algorithms, Classification algorithms ,Neural Net, SVM, Naïve Bayes, BFT, Decision Stump. Correlation investigation of calculations is especially required before actualizing them for the needs of any association. The correlations of calculations are depending on the different parameters, for example, information recurrence, sorts of information and relationship among the attributes in a given informational index. There are number of learning and classifications algorithms are utilized to investigate, learn designs and order information are accessible. Nevertheless, the issue is the one to locate the best algorithm according to the issue and wanted yield. The ideal outcome has dependably been higher precision in foreseeing future qualities or occasions from the given dataset. Calculations taken for the correlations consider are Neural net, SVM, Naïve Bayes, BFT and Decision stump.[6]

Objectives

1. Primary data collection:In our research, we will use Big Data Analytics/Data Mining tools for the primary data collection from various job portals and social media posts with in the Oman.
2. Secondary data collection:We will collect secondary data from the concern organization/Universities/College of Technologies/Industries of Oman.
3. Data Analytics: In our research, we will use Big Data Analytics/Data Mining tools to find Job availabilities based on the individual skill sets. We will use Big Data Analytics/Data Mining tools to find either the available graduates are having enough skills as per our findings.
4. Machine Learning Analytics: In our research, we will use Machine-learning analytics “Apache Spark MLlib/appropriate machine learning tool” to forecast the future Job requirement and Omani graduates availability based on their skill set.

Recommend the educational institution of Oman to add the required training to enhance those skills set in their curriculum based on analytics result, awareness session can be organized in the educational institution for the students.
5. Data Visualization & final report: In our research, we will use data visualization tools to display the analytics results and will compile the final report.

Research Methodology [Describe your Implementation Plan,Time-line and Milestones]

Big Data Analytics/Data Mining will use the following phases of the CRISP Methodology:
Business understanding
Data understanding
Data preparation
Modeling
Evaluation
Deployment
Implementation Plan,Time-line
1. Data collection will take place by using Rapid Miner/Apache Flume from the social media portals or job portal sites.
Time limit: 2 Month
2. Existing data will be collected from the concern authority/universities/college of Technologies/Industries of Oman.
Time limit: 3 Months
3. Big Data Analytics/Data Mining will use current job trend analysis to perform in order to find out the required skills in the available Omani graduates.
Time limit: 2 Months
4. Predictive analysis will perform Machine Learning Analytics to forecast the future Job requirement and candidate availability based on their skill set.
Time limit: 2 Months
5. Results will visualize by using data mining tools in order to prepare and compile the final report.
Time limit: 3 Months

Benefits to Oman

-After the completion of this research author would be able to find the reason behind the unemployment among the Omani graduates. We will submit our findings to the concern department/organization/authority of Oman.
-After the completion of this research author would be able to suggest the required skill set among the Omani graduates to the college of technologies, Universities and other academic institution of Oman.
-After the completion of this research author would be able to forecast job requirement and required job skills, which will guide the Omani students to select appropriate course/training etc.
-Overall, our findings will help to minimize the unemployment in Oman.

Academic, Scientific and/or Innovation Significance

-Impact of proposed research will help to hone the demanding skills in the future graduates.
-Proposed research work will help the academic institution to include required skill set in their curriculum.
-Employment prediction will help to minimize the shortage of skilled graduates in the Oman.

Is this project going to result in a patent?

No

Patent Review (e.g. any previous similar patents in literature, the potential of this project to result in a patent ...)

References

1. https://tradingeconomics.com/oman/unemployment-rate, 03/09/2019
2. M.K. Pusala · J.R. Katukuri · V. Raghavan (✉) Center of Advanced Computer Studies (CACS), University of Louisiana Lafayette, Lafayette, LA 70503, USA e-mail: vijay@cacs.louisiana.edu M.K. Pusala e-mail: mxp6168@cacs.louisiana.edu J.R. Katukuri e-mail: jaykatukuri@gmail.com M. Amini Salehi School of Computing and Informatics, University of Louisiana Lafayette, Lafayette, LA 70503, USA e-mail: amini@louisiana.edu Y. Xie Department of Computer Science, Kennesaw State University, Kennesaw, GA 30144, USA e-mail: yxie2@kennesaw.edu © Springer India 2016 S. Pyne et al. (eds.), Big Data Analytics, DOI 10.1007/978-81-322-3628-3_2
3. " An Overview of Machine learning and its applications", published by Annina Simon1 , Mahima Singh Deo 1,in year 2015: International Journal of Electrical Sciences & Engineering (IJESE) .
4. "Practical Techniques for Interpreting Machine Learning Models: Introductory Open Source Examples Using Python, H2O, and XGBoost", published by Patrick Hall, Navdeep Gill, Mark Chan in February 3,2018:in Engineering research papers: https://www.engpaper.com
5. "Comparative Study of Data Classifiers Using Rapidminer",published by Abhishek Kori in International Journal of Engineering Development and Research (www.ijedr.org) , 2017| Volume 5, Issue 2 | ISSN: 2321-9939
6. "ANALYSIS AND COMPARISON STUDY OF DATA MINING ALGORITHMS USING RAPIDMINER", published by Thirunavukkarasu K,Dr.Manoj Wadhawa in International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.6, No.1, February 2016.