Journal of Science and Sustainable Development (JSSD), 2024, 12(1), 81-87

ISSN: 2304-2702 (print); 2414-4479 (online)
DOI: https://doi.org/10.20372/au.jssd.12.1.2024.0496

RESEARCH PAPER

Job Market for Data Science and Big Data in East Africa

Belachew Regane

Department of Information Technology, Ambo University, Ambo, Ethiopia
Corresponding Author:** Email: **belachew.regane@gmail.com

Abstract

Nowadays big data have been impacting the working culture and working environment of business organizations worldwide. Big data has become an important asset of the business. The main reason is that the insights have been obtained from these data to enable organizations to be better competent in the market. Thus, those organizations using big data for informed decision-making are better benefited than the organizations that are not using it. However, the job role of data science and big data as well as the skills required for the position is not well-identified and distinguished. These issues are the main current challenges in the job market for big data and data science to get the right skilled manpower for the required position. This raises the need for appropriate categories of job positions and skills required to use the data effectively and efficiently for smart decisions and strategic planning in business organizations. The purpose of this study is two-fold. The first is to identify the job market/job position for data science and big data professions. The second is to list out skills required for these job positions in the job market. Data that relate to big data and data science from job advertising posts are collected from online job-related websites and categorized into clusters of job positions. The clustering task is based on the analysis of some commonalities as well as differences observed in the professions. Besides skills required have been identified and listed. The findings show that data scientists, data analysts, and data engineers are identified as popular career paths for the required job market with the required skills in East Africa. As a concluding remark, this research underscores the importance of establishing appropriate job categories and skill sets to effectively leverage big data for informed decision-making and strategic planning within business organizations.

Keywords: Big data, data science, skills required, job positions, job market, job advertisements

Introduction

The advent of big data technology and the data science knowledge domain increased career opportunities to grow exponentially. To fill these needs, Universities and training centers should deliver relevant knowledge and should enable the trainee to possess the required skills (IE-School of Human Science and Technology, 2020). However, there are no clear distinctions between the knowledge domain and skills required for those career opportunities in the job market so hiring companies cannot get the right human resources who possess the right skills that can fit the right position. Davenport and Patil (2012) in Harvard Business Review reported that “if capitalizing on big data depends on hiring scarce data scientists, then the challenge for managers is to learn how to identify that talent, attract it to an enterprise, and make it productive” (Akter *et al*., 2019; Khanra *et al*., 2020).

To identify the knowledge domain and skills required in the field, it is better to understand the relationship between big data, data science, and machine learning. NIST (2019) defines big data as " consists of extensive datasets - primarily in the characteristics of volume, velocity, variety, and/or variability - that require a scalable architecture for efficient storage, manipulation, and analysis."

Most experts and scientists in the discipline of data science define big data by its characteristics. The main characteristics are volume, velocity, and variety. They call it 3V's. However, to define big data in a better way, others include additional characteristics such as veracity, value, vision, verification, validation, complexity, and immutability. Measurement of data in terms of gigabyte, terabyte, or petabyte refers to volume; whereas the speed in which data is generated in terms of time refers to velocity and heterogeneous types of data refer to variety. Besides these main characteristics of Big Data, the added ones have been seen as follows. Vision is used to express the purpose, while verification is used to express confirmation of processed data to some defined specifications. Validation is also used to express the purpose of predefined fulfillment, while the value is used to express that extracted relevant and significant information can be used for a different purpose. Complexity is also another characteristic used to express the difficulty of big data to organize and analyze. Finally, immutability is used to express well-managed big data as it can stay permanent (Oussous *et al*., 2018; Beręsewicz *et al*., 2018).

According to NIST (2019), "Data science is the methodology for the synthesis of useful knowledge directly from data through a process of discovery or hypothesis formulation and hypothesis testing." Since data science includes analytics steps in its process, it is highly linked and used in the analysis of big data. Data science is an interdisciplinary field (as it has been seen in figure 1) so it correlates with statistics, mathematics, data mining, software, and system engineering, algorithms, analytic systems, machine learning, and so on.

Figure 1: Data Science Sub-disciplines (source: NIST (2019))

From these sub-disciplines knowledge domain, machine learning has the lion's share in data analytics to get insight from data of business organizations. Machine learning is one of the areas of computer science. It has been there for decades solving problems in the domain. Arthur Lee Samuel was the first man who defined machine learning as one of the disciplines that were able to learn programs around 60 years ago. The first machine learning algorithms were written by SAS in 1979. Advancements in the area came in the 1990s with approaches including neural networks, decision trees, and ensemble models that integrate multiple machine learning algorithms to enhance the accuracy of predictions. Today,

the most popular online service provider like Google, Amazon, and Netflix are using machine learning in their work. There are two types of machine learning (Muezzinoglu *et al*., 2018).

Currently, data scientists are commonly using machine learning in big data analytics. It is used in the application areas such as the financial industry, aviation industry, crime protection, and identifications, robotics industry, telecommunication industry, autonomous cars industry, and others to analyze data, predict, and make intelligent decisions (Muezzinoglu *et al*., 2018; Strau, 2018).

In machine learning, the model-based approach discussed by Breuker (2014) and Muezzinoglu et al., (2018), is one of the techniques implemented in data analytics. By choosing one of the available algorithms in the library, tasks will overtake and be executed on the model. This enables us to have algorithms for the specific implementation.

In East Africa, countries like Kenya, Tanzania, Uganda, Rwanda, and Ethiopia were witnessing a rise in opportunities for data professionals. Industries such as finance, telecommunications, healthcare, agriculture, and e-commerce were actively seeking skilled individuals to help them make sense of their data and gain a competitive edge (KICTANet, 2023).

Here, the question is why a need to know the role and skills required in data science and big data. The basic reason is to achieve the objective of the study. That is to identify the job categories and skills required in the area so that business companies hire the right human resources they need. Besides, University and educational training centers deliver relevant knowledge and skills to fill the needs of the industries. Therefore, to achieve the objective of the study, the following research questions have been developed.

1. What are the job market/job positions for data science and big data professionals in East Africa so that hiring companies get the right human resources that can fit the right position?

2. What are the skills required in the job market/job positions of data science and big data professionals in East Africa so that hiring companies get the right human resource that possesses the right skills?

Methods

The approach used for this study is a systematic review and cluster analysis method is used to analyze the collected data. The following search keys were developed to collect relevant primary data based on the research questions.

((Careers in Data Science OR Job Roles in Data Science) AND (Data Analyst OR Data Engineer OR Data Scientist OR Database Administrator OR Machine Learning Engineer OR Data Architect OR Business Analyst OR Data and Analytics Manager OR Data Visualization Specialist OR Consultant Data Management OR Big Data Engineer OR Data Science Specialist OR Research Data Manager OR Data Lifecycle Expert OR Analytics Lead OR Data Science and Big Data Consultant OR Big Data Specialist OR Dig data software engineer OR Lead Data Engineer OR Data Ops Engineer OR Data Management Specialist OR Lead Data Solutions Architect OR Data Protection and Storage Engineer OR Systems Data Protection and Storage Engineer OR Machine Learning Software Engineer)).

To search and collect relevant data, the following Websites and search engines are used as sources of data. These websites are selected based on their relevance and popularity in posting job vacancies in East Africa.

1. https://www.EthioJobs.net
2. https://www.LinkedIn.com
3. https://www.tapwage.com
4. https://www.unjobnet.org
5.ttps://www.ethiopiawork.com/africawork.com
6. https://www.Google.com/Jobs vacancy

The following Inclusion and Exclusion criteria are used to filter collected data for relevance. As inclusion criteria, any industry with relevant data concerning data science, and big data is

included. The data should be only from East African countries such as Djibouti, Eritrea, Ethiopia, the Democratic Republic of the Congo, Somalia, Burundi, Kenya, Rwanda, South Sudan, Uganda, and Tanzania. The years of vacancy posted should be between 2012 and 2023. Whereas job position with the same name but which has no link with data science and big data are excluded from collected data. Also, skills posted by job vacancy advertising but not related to data are excluded from skills required for job position categories.

Results and discussions

From many posted job advertising vacancies, 22 job vacancies that relate to big data and data science are collected. According to IE-School of Human Science and Technology (2020) nowadays, a few career paths in big data and data science in a worldwide context are Machine learning engineer, Data architect, Statistician, Data analyst, Chief technology officer, chief data officer (CDO), Application Architect, Project manager, Market research analyst, Business/analytics translator, Data scientist, and so on. Based on this fact, the collected data was analyzed and categorized into different groups on their similarities and differences using the cluster analysis method. Furthermore, based on the inclusion and exclusion criteria two vacancies are excluded so that the numbers are reduced to 20. Finally, from 20 vacancies available for grouping into their similarities, 5 clusters have been formed. 8 vacancies are categorized to data scientist, 8 vacancies are categorized to data analyst, 2 vacancies are categorized to data engineer, 1 vacancy is categorized to data management, and 1 vacancy is categorized to big data program officer.

Intern and advisory/consultancy are categorized under relevant job vacancies, not separately. The result revealed that data scientists and data analysts are job positions in data science and big data professions that most companies in East Africa need to hire. The related skills required for these job positions are shown in the tables below (see Tables 2, 3, and 4 for details).

Table 1. Clusters of job vacancy

SN	Job vacancy	Frequency	%
1	Data Scientist	8	40
2	Data Analyst	8	40
3	Data Engineer	2	10
4	Data Management	1	5
5	Big data program officer	1	5

The findings show that data scientist and data analyst are the two job positions in data science and big data profession that most companies in East Africa need to hire for their data-related tasks. On the other hand, the educational qualification for big data and data science job market has seen almost the same in all job role categories. The skills required are also overlapping to some extent but have important differences. This implies that the skills difference may arise from exposure to the industries or skills acquired through different skills training centers.

Existing related literature works also highlights the increasing importance of big data in business organizations and the challenges and opportunities in the job market for data science and big data professions. This research emphasizes the need for appropriate job categories and skills for the effective use of big data, particularly in East Africa, while works of literature focuses on the analysis of job advertisements and the multi-faceted nature of big data job skills, including the value placed on soft skills (Gardiner et al., 2018; Regane *et al*., 2024).

Table 2. Skills and educational qualifications required for the role of data scientist

Roles	Skills required	Educational Qualifications
Data scientist	Statistical analysis and computer programming Deep analytical and research skills Personal organization and planning skills Manage and analyze big data Analyze and evaluate the data Create new models of analysis Assist the users Perform the technology watch Quantitative analytics or data modeling Deep understanding of predictive modeling, machine learning, clustering and classification techniques, and algorithms Fluency in a programming language (Python, C, C++, Java, SQL) Familiarity with Big Data frameworks and visualization tools (Cassandra, Hadoop, Spark, Tablea	Data science, Mathematics, Statistics, Computer science, Software engineering, Systems engineering and Information systems

As claimed by NIST (2019) "A data scientist is a practitioner who has sufficient knowledge in the overlapping regimes of business needs, domain knowledge, analytical skills, and software and systems engineering to manage the end-to-end data processes in the analytics life cycle."

Table 3. Skills and educational qualifications required for the role of data analyst

Roles	Skills required	Educational Qualifications
Data analysts	Experience in data models and reporting packages Ability to analyze large data sets An analytical mind and inclination for problem-solving Deep understanding of predictive modeling, machine learning, clustering and classification techniques, and algorithms Knowledge of statistics and experience using statistical packages for analyzing datasets (Excel, SPSS, SAS, etc.) Familiarity with big data frameworks and visualization tools (Cassandra, Hadoop, Spark, Tableau Coding database environments in MySQL, SQL, or other DB languages Fluency in a programming language (Python, C, C++, Java, SQL) Programming experience with frameworks including XML, Javascript, and ETL	Data science, Statistics, Information Technology, Mathematics, Computer Science, Economics, and Information management

Modernanalyst.com (2020) defines a data analyst as "The Data Analyst is the professional whose focus of analysis and problem-solving relates to data, types of data, and relationships among data elements within a business system or IT system." On the other hand, the role of data engineers is defined by Furbush (2018) as "Data engineers are concerned with the production readiness of that data and all that comes with it: formats, scaling, resilience, security, and more".

Table 4: Skills and educational qualifications required for the role of the data engineer

Roles	Skills required	Educational Qualifications
Data Engineer	Experience doing quantitative analysis Fluency in SQL or other programming languages. Basic understanding of statistical analysis Strong coordination and project management skills Excellent analytic skills Expertise in designing and maintaining databases (object, columnar, in-memory, relational) Proven track record of successful communication of data infrastructure, data models, and data engineering Experience with relational data stores as well as one or more NoSQL data stores (e.g. Mongo, Cassandra) Prior experience in data warehouse modernization building complete data warehouse solutions, star/snowflake schema designs, infrastructure components, ETL/ELT pipelines, and reporting/analytic tools Experience building production-grade data backup/restore, and disaster recovery solutions Hands-on experience with batch and streaming data (e.g., Cloud Dataflow, Beam, Spark, Cloud Pub/Sub, Apache Kafka) Advanced SQL skills, and proficiency in one or more programming languages such as Python Familiarity with Python data science tooling (pandas, scipy, sklearn) Demonstrated proficiency with data structures, algorithms, distributed computing, and storage systems	Computer Science, Mathematic, Physics, Engineering, Statistics or other technical fields

Conclusion

To cope with the changing environment in business organizations due to emerging technologies like big data, it is important to understand the knowledge domain and skills required in data science and big data. To identify professions' knowledge domain and skills required in the area, posts on job advertising websites were assessed and analyzed. Finally, data are clustered into different categories based on their similarities. The result of the study shows that data scientists, data analysts, and data engineers are the leading job positions of big data and data science professionals in East Africa. Besides, commonly required skills for these job positions in the professions are listed. Identifying these professional job positions and the skills required helps business organizations

hire the right person for the right position to be competent and profitable.

Acknowledgments

I would like to express my deepest gratitude to Dr. Sarah Beeckam and Dr. Norah Power for their exceptional supervision and invaluable guidance throughout this research. Their expertise, support, and insightful feedback have been instrumental in shaping this study and have greatly contributed to its successful completion. I am truly grateful for their mentorship and unwavering commitment to excellence.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Akter, S., Fosso Wamba, S., Barrett, M., and Biswas, K. 2019. How talent capability can shape service analytics capability in the big data environment? Journal of Strategic Marketing, 27(6), 521-539.

Beręsewicz, M. Lehtonen, R., Reis, F., Di Consiglio, L. and Karlberg, M. 2018. An overview of methods for treating selectivity in big data sources, Eurostat -2018 edition.

Breuker, D. 201). Towards Model-Driven Engineering for Big Data Analytics – An Exploratory Analysis of Domain-Specific Languages for Machine Learning, ResearchGate.

Davenport, T. H. and Patil, D.J. 2012. Data Scientist: The Sexiest Job of the 21st Century, available [online] hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century, From the October 2012 Issue

Furbush, J. 201). Data engineering: A quick and simple definition, available [online] https://www.oreilly.com/content/data-engineering-a-quick-and-simple-definition/

Gardiner, A., Aasheim, C., Rutner, P., and Williams, S. 2018. Skill requirements in big data: A content analysis of job advertisements. Journal of Computer Information Systems, 58(4), 374-384

IE-School of human science and technology, 2020. Master in Business Analytics & Big Data: Drive business transformation through Big Data, Artificial Intelligence, and Business Analytics available [online] https://www.ie.edu/school-human-sciences-technology/masters/master-business-analytics-big-data/high-impact-careers/

Khanra, S., Dhir, A., and Mäntymäki, M. 2020. Big data analytics and enterprises: a bibliometric synthesis of the literature. Enterprise Information Systems, 14 (6), 737-768.

KICTANet, 2023. The Rise of Tech-Driven Solutions in East Africa: Trends and Opportunities, available [online], https://www.linkedin.com/pulse/rise-tech-driven-solutions-east-africa-trends-opportunities-a4wrf/

Modern analyst.com, 2020. The Data Analyst Role: High-level definition of the Data Analyst role, available [online], modernanalyst.com/TheProfession/Roles/DataAnalyst/tabid/189/Default.aspx

Muezzinoglu, K., Suplee, C. and Stewart, D. 2018. Machine Learning Use Cases in Financial Crimes: Ten practical and achievable ways to put machine learning to work, SAS Institute Inc.

NIST, 2019. Big Data Interoperability Framework: Volume 1, Definitions, Version 3, available [online] https://doi.org/10.6028/NIST.SP.1500-1r2

Oussous, A., Benjelloun, F., Lahcen, A. A. and Belfkih, S. 2018. Big Data Technologies: A survey, Journal of King Saud University – Computer and Information Sciences 30 (2018) 431–448.

Regane, B. and Power, N., Beecham, S. and Lemma, D. 2024. Stakeholder Requirements for Big Data Projects: A Multi-Case Study Review. Available at SSRN: https://ssrn.com/abstract=4768913 or http://dx.doi.org/10.2139/ssrn.4768913

Strau, S. 2018. From Big Data to Deep Learning: A Leap Towards Strong AI or 'Intelligentsia Obscura'? big data and cognitive computing.