International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 7 Issue 11, November - 2023, Pages: 1-6

Integrating Big Data for Evaluating the Impact of Curriculum Deficiency on Labour Market Preparedness

Akinode John Lerkan1 and Bada Olufunke2

1Dept. Of Computer Science, Federal Polytechnic Ilaro, PMB 50,Ilaro,Ogun State,Nigeria
2Dept. Of Computer Science, Federal Polytechnic Ilaro, PMB 50,Ilaro,Ogun State,Nigeria

Abstract*—This research addresses the pressing issue of skills mismatch in Nigeria's labor market, particularly affecting graduates in science and technology fields. With youth unemployment rates ranging from 11.7% to 36.5%, it is imperative to bridge the gap between educational curricula and industry demands. This study proposed an innovative tool to perform labor market demand analysis and curriculum gap assessments by utilizing blockchain, artificial intelligence, and big data analytics. By collecting and analyzing data from diverse job portals, the tool aims to provide real-time insights into skills required by employers. Additionally, blockchain technology ensures secure and verifiable credentials for graduates. This research seeks to empower graduates and educational institutions, contributing to sustainable economic growth in Nigeria.*

Keywords: Artificial Intelligence (AI),Skills mismatch, Labor market, Nigeria, AI tool, curriculum assessment, Blockchain Technology, Big data Analytics,

1. INTRODUCTION

In this rapidly evolving landscape, it becomes imperative for Nigeria to tackle the complex issues posed by the digitization of society. The push towards a more technologically advanced society offers a myriad of opportunities, yet it also ushers in a new era of global competition and the automation of tasks previously carried out by human labor (Ketamo H. & Anu P.R., 2019). Consequently, companies in Nigeria are increasingly vocal about the dearth of skilled workers and the pressing issue of global skill mismatches.The scenario is particularly grim for the nation's graduates. Nigerian universities continue to produce a substantial number of graduates annually, many of whom find themselves entering the labor market with little hope of securing their desired positions. The problem of unemployment among Nigerian youths, especially those with a background in science and technology, has reached alarming proportions (Adeyemo S.A. et al, 2010). This situation not only poses a challenge to the individual aspirations of these graduates but also hinders the overall economic growth and the nation's pursuit of the Sustainable Development Goals by 2030 (Olojuolawe, R. S et al, 2019).

Disturbingly, the issue of youth unemployment has remained consistently high, hovering between 11.7% and a staggering 36.5% from 2017 to 2018 (Olojuolawe, R. S et al, 2019). According to Okeke Udoka Emmanuel (2020), this increase is explained by the widening gap between graduates' skill sets and the real needs of the job market. The Nigeria Job Report of 2015 delves deeper into the systemic issues, asserting that the problems in Nigeria's skill-building system originate from both the supply and demand sides. It highlights inefficiencies in the mechanisms and institutions designed to facilitate the transition from education to the workforce, underscoring the need for a comprehensive method to tackle these challenges.

The research is focused on designing an innovative AI tool capable of conducting labor market demand analysis and curriculum gap assessments for selected institutions in Nigeria. Additionally, it aims to establish a comprehensive job classification taxonomy framework for the entire nation.

1. Statement of the Problem

The prevalence asymmetric information in the Labour market have been on the both side of the market. Prospective employers have vague knowledge about the available workers and their attributes. Also, potential employee is uncertain whether their dream company would make a good employer.The gap between the skills that higher education students gain and the abilities that businesses actually want in the workforce results in glaring shortages of the necessary talents. This is a serious obstacle to Nigeria's growth and development and has rendered many Nigerian graduates unemployable.Making sure the curriculum is current and uses terminology from the labor market are the two main obstacles that most higher education institutions encounter when creating a curriculum related to the labor market. Since most curricula are written in academic languages, they frequently omit terms that are used in the workplace. This can make it difficult for students to apply the skills they learn in the classroom to real-world situations.

2. Objectives of the Research

The primary aim of this research is to create an AI tool capable of performing an analysis of labor market demand and assessing curriculum gaps for specific Universities and Polytechnics in Nigeria.. This study would focus on ICT, Accounting and Electrical

Electronics industries. . It will teach participants how artificial intelligence (AI), blockchain technology, and data analytics can be leveraged to add value to higher education institutions' curriculum revamps in order to satisfy employers' demands and close the gap between supply and demand in the labor market.

The specific objectives of this work include:

i. To Collect online job advertisement data from three publicly openly machine -readable online job portals in Nigeria between 2019-2023.
ii. Develop a big data tool to create and visualize skill map for the selected industries.
iii. Machine –read and analyses the curricula of selected institution in Nigeria.
iv. Create an AI tool and algorithms designed to perform a gap analysis between the labor market and curricula, with the goal of identifying skills deficiencies. Develop a blockchain technology to obtain a non-immutable credentials of the potential graduates in case of occupational mobility.
v. Presentation of analysed data for better decision making

3. Artificial Intelligence

Artificial Intelligence,or AI, denotes the evolution of computer systems capable of executing tasks traditionally linked with human intelligence. These tasks encompass abilities like speech recognition, decision-making, visual perception, and language translation. AI systems enhance their proficiency by assimilating information and detecting fundamental patterns (Russel & Norvig, 2016).

3.1 **Big Data**

Massive datasets that can be computationally analyzed to find patterns, trends, and correlations are referred to as "big data." Three main components are involved: variety (different types of data), velocity (high speed at which data is generated), and volume (large amounts of data). Numerous fields use big data for insights, forecasting, and decision-making (Manyinka et al., 2011).

3.2 Blockchain

Blockchain is a decentralized and distributed ledger technology that securely records transactions across numerous computers, preserving the integrity and security of the data. Each block within the chain comprises multiple transactions, and once recorded, the information in a block remains unchangeable retroactively. Its prominent application is as the underlying technology for cryptocurrencies such as Bitcoin (Narayan, 2016).

4.3 Labour market Intelligence

Labour Market Intelligence (LMI) is a burgeoning interdisciplinary field that is gaining momentum in both academia and industry. It primarily involves extracting valuable insights from unstructured text, like curricula and job postings, with a focus on streamlining e-recruitment processes. This is achieved by automating resume management and matching candidate profiles to job descriptions. Additionally, there is a pressing need to automate activities within human resources departments (Mario Mezzanzanica et al., 2019).LMI encompasses the utilization and development of AI algorithms and frameworks to analyze Labour Market Data, thereby aiding decision-making. The surge in online job postings offers a prime opportunity for real-time monitoring of the labour market (Mezzanzanica, Mercorio et al., 2017). According to Mario Mezzanzanica et al. (2019), the demand for skills and qualifications in emerging labour markets has undergone significant shifts, along with a notable transformation in the skill requirements for new jobs. Leveraging big data ensures that information is systematically collected, organized, and analyzed for immediate monitoring and analysis of labour market dynamics and trends. This grants a competitive edge over traditional administrative and survey-based data sources, which may take several months to compile. Real-time analysis of the labour market offers swift insights into market demands, supplying valuable information about sought-after skills and specific job requirements that are not typically covered in traditional surveys.Artificial intelligence and Big data analytics can provide a new method to determine competence needs in a workplace (Ketamo H et al,2019). Industries are constantly searching for new employee through web resources (Roberto Boselli et al,2017). The humongous nature of the data shared via the digital channel create an open data that can be used by curriculum designers. The past research depicts that big data can assist companies to prepare for future trends and it uses large volume of data from which analyst extract patterns. A higher percentage of company data is unstructured; this make analytics more difficult than in analyzing structured data

4. Related works

In their 2017 work, Mezzanzanica, Mercorio, et al. introduced an innovative technique for automatically categorizing web job listings according to a standardized taxonomy of occupations. Their approach employed machine learning methods for text classification, utilizing both Bag of Words and Word2Vec feature extraction techniques. The researchers created a practical data system for European call for tender projects granted by EU organizations, focusing on classifying job vacancies using machine learning algorithms. They meticulously designed and assessed multiple classification pipelines to assign ISCO occupation codes to diverse job listings in English.The establishment of a classification taxonomy is crucial for scrutinizing, exchanging, and comparing dynamics within the web labor market. One prominent example of a standardized classifier is the ISCO hierarchical classifier, facilitated by the European Network on Regional Labour Market Monitoring. Over time, a handful of commercial skill-matching products have been developed, including Burning Glass, Workday, Pluralsight, and TextKernel.The EU and Eurostat conceived the ESSnet Big Data project in 2016 with the aim of seamlessly integrating big data into the routine production of official statistics. This involves exploring selected Big Data sources through pilot projects and constructing practical applications. Similarly, the EU's CEDFOP agency initiated a call-for-tender to create a system capable of collecting and categorizing web job listings from EU countries. The principal objective of this project is to transform data extracted from web job listings into actionable knowledge for policy formulation and improved decision-making.
(Harri Ketamo et al,2019) adopted Natural Language processing,cognitive AI and big data to create a real time understanding of skills, competencies, knowledge and abilities that are sought for in workplace. They analysed an open big data by using a custom software (Head AI platform) to create information about expected competencies stated in job adverts online. They used Natural language processing(NLP) algorithms to build micro level understanding on competencies in real time for workplace in Finland.

In the study conducted by Roberto Boselli et al. (2017), they introduced WoLMIS, a system designed to aggregate job postings from various European web job platforms and categorize them according to the international taxonomy of occupations. The initiative was spearheaded by Cedefop, a European agency dedicated to promoting the advancement of European Vocational Education and Training. Since June 2016, WoLMIS has been operational on the Cedefop data center, systematically gathering and classifying job vacancies.WoLMIS fulfills two primary functions: firstly, it collects job vacancies by scraping selected countries' internet-based job portals. Secondly, it employs Machine Learning algorithms for text classification to assign each job vacancy a classification based on the international standard classification of occupations (ISCO) taxonomy. The process of classifying and analyzing web-based job vacancies and occupations has presented a notable challenge (Roberto Boselli et al., 2017).Analyzing web job advertisements encounters two technical hurdles: firstly, data obtained from online job portals is in various languages and adheres to diverse taxonomies. Secondly, job adverts are predominantly presented in an unstructured format and published as plain text (Roberto Boselli et al., 2017). Consequently, many job placement authors employ job titles and terminology that do not align with standardized classification systems.To address the need for a unified label classifier capable of recognizing ISCO codes from textual job vacancies, the WoLMIS system employs supervised learning. Various machine learning algorithms were selected and implemented in WoLMIS, and subsequently evaluated using a set of English job postings. The employed methods include Linear Support Vector Machines (SVM), Support Vector Machines with Radial Basis Function (RBF) Kernel, Random Forests (RF), and Artificial Neural Networks (ANN). In the preliminary prototype, linear SVMs are coupled with a bag-of-words technique for feature extraction.
In their work, Zhu et al. (2016) created a database of online job openings sourced from Chinese websites between 2014 and 2015. They then employed topic recognition algorithms to discover latent subjects related to the recruitment market and track changes in these topics over time. The study involved the extraction of job openings from Chinese websites, which were then categorized based on the CGCO (People's Republic of China Grand Classification of Occupations).

Xu et al. (2017) implemented both SVM and Neural Network Classifiers on their dataset. A considerable amount of commercial skill matching tools have been developed in the past, for example, Burning Glass, Workday, Pluralsight and Text kernel. Also, the Google Cloud Job API provides a service for classifying job vacancies and skill identification leveraging on a standard taxonomy.

5 Methodology

5.1 Materials

The proposed dataset for this research will be from scrapping ten (10) different Nigerian job websites. We were able to scrape 500 job opportunities each posted on 10 different job portals. The scraped data was structured in a tabular format with features like “Job title,Name of the Company, Job Description, Job category, Experience level e.t.c). However, we will employed some research

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 7 Issue 11, November - 2023, Pages: 1-6

assistants to help in labeling of all job descriptions based on the benchmark of the International Labour Organisation (ILO) . The labelled data will serve as a target variable for the modelling using Machine learning approach.

Algorithm:

Data Collection
- Scrape data from ten (10) different Nigerian job websites.
- Obtain 500 job opportunities from each of the ten job portals.
- Structure the scraped data in a tabular format with features:
  - \(\text{JobTitle}_i, \text{CompanyName}_i, \text{JobDescription}_i, \text{JobCategory}_i, \text{ExperienceLevel}_i, \ldots \) where \(i = 1, 2, \ldots, 500 \times 10 \).

- Employ research assistants for labeling job descriptions based on ILO benchmarks.

5.2 Text Processing

Natural language toolkit (NLTK) which is a powerful tool for preprocessing text data was adopted for the text processing on Python 3.7. We will be able to apply tokenization, removing special characters, dealing with emojis, removing of HTML tags, removal of punctuations, changing all characters to lowercase and removal of unnecessary words which are not fit for the modelling.

The feature extraction process involved the application of Term-frequency-inverse document frequency (TF-IDF), a statistical metric assessing the relevance of a word within a document set. This was determined by considering both the frequency of a word's occurrence in a specific document and its inverse frequency across the entire document collection.

Algorithm:

- Use Natural Language Toolkit (NLTK) for text preprocessing in Python 3.7.
  - Tokenization: \(\text{Tokenize}(\text{JobDescription}_i) \)
  - Removal of special characters: \(\text{RemoveSpecialChars}(\text{TokenizedText}) \)
  - Handling of emojis: \(\text{HandleEmojis}(\text{Text}) \)
  - Removal of HTML tags: \(\text{RemoveHTMLTags}(\text{Text}) \)
  - Removal of punctuations: \(\text{RemovePunctuations}(\text{Text}) \)
  - Convert all characters to lowercase: \(\text{ToLowercase}(\text{Text}) \)
  - Removal of unnecessary words: \(\text{RemoveStopWords}(\text{Text}) \) (not fit for modeling).

- Apply Term-frequency-inverse document frequency (TF-IDF) for feature extraction:
  - \(\text{TF-IDF}(w, d) = \text{TF}(w, d) \times \log\left(\frac{N}{\text{DF}(w)}\right) \)
    - \(\text{TF}(w, d) \) is the term frequency of word \(w \) in document \(d \).
    - \(\text{DF}(w) \) is the document frequency of word \(w \).
    - \(N \) is the total number of documents.

5.3 Modelling

Based on this research, we propose to apply three machine learning models which are Support vector classifier, Naive bayes and Neural Networks. We will carry out a Gridsearch of 5fold cross validation using multiple hyperparameters for each model such that Gridsearch will look for the best hyperparameter and our evaluation metrics were computed based on the analysis. Prior to this, due to the number of features generated based on the feature extraction method, we notice that there might be some redundant features which are not necessarily useful for the modelling. Therefore, we applied a pre-trained dimensionality reduction approach “Word2Vec”. The best model was selected based using some major classification evaluation metrics. For the purpose of this research, we used F1 Score, Precision, Recall and Classification Accuracy.

Algorithm:

- Apply three machine learning models:
    - Support Vector Classifier (SVC)
    - Naive Bayes (NB)
    - Neural Networks (NN)

- Perform Gridsearch of 5-fold cross validation with multiple hyperparameters for each model:
    - Find the best hyperparameters.

- Apply a pre-trained dimensionality reduction approach "Word2Vec" to handle redundant features.

- Evaluate models using major classification evaluation metrics:
    - \(\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \)
    - \(\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \)
    - \(\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \)
    - \(\text{Classification Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} \)

- Select the best model based on evaluation metrics.

6 Conclusion

In conclusion, this research addresses a critical issue in Nigeria's evolving digital landscape: the mismatch between the skills acquired by graduates and the demands of the labor market. The persistently high levels of youth unemployment underscore the urgency of this problem.Utilizing cutting-edge technologies like AI, blockchain, and advanced data analytics, this research introduces a holistic approach to address this disparity.The proposed AI Tool aims to provide valuable insights into labor market dynamics, offering a real-time understanding of skills and competencies sought after by employers. This tool's potential impact is significant, not only for graduates seeking meaningful employment but also for educational institutions striving to align their curricula with industry needs.Furthermore, the integration of blockchain technology in securing graduates' credentials for potential occupational mobility is an innovative and forward-thinking approach. This ensures the authenticity and immutability of qualifications, bolstering trust in the labor market.By collecting and analyzing data from various job portals, this research endeavors to create a comprehensive framework that not only identifies skills gaps but also visualizes them, providing a tangible representation of the challenges at hand. This visual representation will serve as a powerful tool for decision-makers in academia and policy.

In essence, this research not only identifies a critical issue but also proposes a technological solution with the potential to revolutionize the way education and employment intersect in Nigeria. By leveraging cutting-edge technologies and data-driven approaches, this work aspires to empower both graduates and educational institutions, ultimately contributing to the nation's sustainable development and economic growth in line with the 2030 Sustainable Development Goals.

References

Adeyemo, S. A., Ogunleye, A. O., Oke, C. O., & Adenle, S. O. (2010). A survey of factors determining the employability of science and technology graduates of polytechnics and universities in the Nigerian labour market. Journal of Science and Technology Education Research, 1(5), 99-106. Available online at http://www.academicjournals.org/JSTER

Boselli, R., Cesarini, M., Marrara, S., & Mercorio, F. (2018). WoLMIS: A labor market intelligence system for classifying web job vacancies. Journal of Intelligent Information Systems, 51(2), 1-26. DOI: 10.1007/s10844-017-0488-x.

Ketamo, H., Moisio, M., Passi-Rauste, A., & Alamäki, A. (2019). Mapping the Future Curriculum: Adopting Artificial Intelligence and Analytics in Forecasting Competence Needs. In Sargiacomo, M. (Ed.), Proceedings of the 10th European Conference on Intangibels and Intellectual Capital ECIIC 2019 (pp. 144-153). Chieti-Pescara, Italy.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Mezzanzanica, M., & Mercorio, F. (2017). Using Machine Learning for Labour Market Intelligence. In ECML PKDD 2017, LNCS. Springer.

Mezzanzanica, M., & Mercorio, F. (2019). Big data enables Labour Market Intelligence. In Encyclopedia of Big Data Technologies. Springer.

Nigeria: Skills for Competitiveness and Employability. (2015). This policy note is a companion volume to the Nigeria jobs report entitled: More, and More Productive, Jobs for Nigeria – A Profile of Work and Workers.

Okeke, U. E. (2020). Skills Mismatch and Dearth of Vocational Education in the Nigerian Labour Market: A Cog in the Wheels of National Development. American Journal of Educational Research, 8(3), 173-181. doi: 10.12691/education-8-3-8.

Xu, H., Gu, C., Zhou, H., Zhang, J. (2017). Hr-ctc: A large human resource corpus for text classification. arXiv preprint arXiv:1705.06123.

Zhu, C., Zhu, H., Xiong, H., Ding, P., Xie, F. (2016). Recruitment market trend analysis with sequential latent variable models. In: Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, pp. 383–392. ACM, New York, NY, USA. DOI: 10.1145/2939672.2939689. URL: http://doi.acm.org.proxy.unimib.it/10.1145/2939672.2939689.

International Labour Organization. (n.d.). International Labour Organization. Retrieved from https://www.ilo.org/global/lang--en/index.htm

Natural Language Toolkit — NLTK 3.5 documentation. (n.d.). Retrieved from https://nltk.org

Tf-idf :: A Single-Page Tutorial - Information Retrieval and Text Mining. (n.d.). Retrieved from https://tfidf.com