American Journal of Engineering and Applied Sciences

Original Research Paper

Propose a System of Collecting and Analyzing Casual Labor Data in Libya Based on Distributed Systems and Weka Visualization

[1]Magda Shuayb, [2]Ashraf Fadiel and [3]Taher Abu-Lebdeh

[1]Department of Civil Engineering, The Higher Higher Institute of S&T, El-Beida, Libya
[2]Department of Civil Engineering, Omar Al-Mukhtar University, El-Beida, Libya
[3]Civil, Architectural, and Environmental Engineering Department, North Carolina A&T State University, Greensboro, NC, USA

Article history
Received: 18-12-2023
Revised: 01-01-2024
Accepted: 03-01-2024

Corresponding Author:
Ashraf Fadiel
Department of Civil Engineering, Omar Al-Mukhtar University, El-Bieda, Libya
Email: ashraf.fadiel@omu.edu.ly

Abstract: The aim of this research is to propose a system for collecting and analyzing casual labor data in Libya, leveraging distributed systems and the Weka visualization tool. The current labor market in Libya relies heavily on casual labor, which poses challenges in terms of data collection and analysis due to its decentralized nature. The proposed system aims to address these challenges by utilizing distributed systems to efficiently gather data from various sources and employing the Weka visualization tool for comprehensive analysis. The system consists of two main components: Data collection and data analysis. For data collection, a distributed system architecture is proposed, which enables the collection of labor data from multiple sources such as job platforms, recruitment agencies and government databases. The distributed system ensures efficient and reliable data retrieval, incorporating mechanisms for data quality control and validation. The proposed system has several advantages. Firstly, it improves the efficiency and accuracy of data collection by leveraging distributed systems, reducing manual efforts and minimizing errors. Secondly, the utilization of Weka for data analysis enhances the understanding of casual labor trends, patterns and demographics, enabling better decision-making in labor market policies and resource allocation. Additionally, the system contributes to the overall digitalization and modernization of labor market data management in Libya.

Keywords: Casual Labor, Libya, Distributed Systems, Weka Visualization

Introduction

The labor market plays a vital role in the socio-economic development of any country. In Libya, the labor market is characterized by a significant presence of casual labor, where individuals engage in short-term or temporary work arrangements (El-Fadil, 2009). This form of employment poses challenges in terms of data collection and analysis due to its decentralized nature and the absence of a centralized labor information system. However, with the advancements in technology and the availability of data analysis tools, there is an opportunity to develop a system that efficiently collects and analyzes casual labor data to gain insights into labor market dynamics (Rado, 2019; Desiere and Costa, 2019; Gary, 2019; Manalu *et al*., 2023). The purpose of this research is to propose a system for collecting and analyzing casual labor data in Libya, leveraging distributed systems and the Weka visualization tool. The proposed system aims to address the existing challenges in data collection and analysis, providing a comprehensive solution to understand the patterns, trends and demographics of casual labor in Libya (Rado, 2019; Desiere and Costa, 2019; Gary, 2019; Manalu *et al*., 2023; Breza *et al*., 2021; Spaargaren, 2020; Bakthavatchalam *et al*., 2022). The need for such a system arises from the importance of having accurate and up-to-date labor market information to inform policy-making, resource allocation and decision-making processes. Currently, the lack of centralized and reliable labor data hinders the ability of stakeholders, including government agencies, policymakers and researchers, to make informed decisions regarding labor market interventions (Rado, 2019; Desiere and Costa, 2019; Gary, 2019; Manalu *et al*., 2023; Breza *et al*., 2021; Spaargaren, 2020; Bakthavatchalam *et al*., 2022). By utilizing distributed

systems, the proposed system enables the collection of labor data from various sources, including job platforms, recruitment agencies and government databases. This distributed approach ensures the efficient retrieval and integration of data, overcoming the challenges posed by the decentralized nature of casual labor. Additionally, the use of the Weka visualization tool provides powerful data analysis capabilities, allowing for the application of data mining and machine learning techniques to gain valuable insights from the collected data (Gary, 2019; Manalu *et al*., 2023; Breza *et al*., 2021; Spaargaren, 2020; Bakthavatchalam *et al*., 2022). The proposed system offers several benefits. Firstly, it improves the efficiency and accuracy of data collection by leveraging distributed systems, enabling stakeholders to access comprehensive and reliable labor market data. Secondly, the use of the Weka visualization tool enhances the understanding of casual labor trends, patterns and demographics through advanced data analysis techniques. This, in turn, facilitates evidence-based decision-making and policy formulation in relation to labor market interventions (Nguyen *et al*., 2019; Bakthavatchalam *et al*., 2022; Hamed *et al*., 2021; Bourechak *et al*., 2023; Heikkonen, 2021). To develop and evaluate the proposed system, a multidisciplinary approach will be adopted, incorporating principles from data science, distributed systems and labor market analysis. The research will involve a thorough literature review, system design and implementation, data collection from various sources and analysis using the Weka visualization tool (Bourechak *et al*., 2023; Heikkonen, 2021). The effectiveness and utility of the system will be evaluated through real-world data, providing insights into the feasibility and applicability of the proposed approach. The proposed system for collecting and analyzing casual labor data in Libya based on distributed systems and Weka visualization offers a promising solution to address the challenges in labor market data management (Nguyen *et al*., 2019; Bakthavatchalam *et al*., 2022; Hamed *et al*., 2021; Bourechak *et al*., 2023; Heikkonen, 2021). By leveraging technology and advanced data analysis techniques, the system aims to provide stakeholders with valuable insights into the dynamics of casual labor, contributing to evidence-based decision-making and policy formulation in the labor market. Work is one of the most important requirements of life in human societies (Bourechak *et al*., 2023; Heikkonen, 2021). Therefore, different societies seek to provide job opportunities for individuals who are able to work in order to provide a decent life for the individual and their families. Libya is a coastal country in the Mediterranean. Many people travel to it in order to work. Most of the people traveling to Libya are immigrants from the neighboring countries of Libya, hoping to improve their lives and provide a decent life for their families (Gomes *et al*., 2019; Bourechak *et al*., 2023; Heikkonen, 2021). This group of people is defined as casual labor. In this research, a proposal is made to collect casual labor data in various Libyan cities using specialized centers and distributed systems and the program is prepared and implemented in the Python language (Nguyen *et al*., 2019; Bakthavatchalam *et al*., 2022; Hamed *et al*., 2021; Bourechak *et al*., 2023; Heikkonen, 2021).

Materials and Methods

The research methodology involves a combination of literature review, system design and implementation, data collection and analysis using Weka. The Waikato environment for Knowledge analysis (Weka) was developed due to the perceived need for a single workbench that would provide researchers simple access to cutting-edge machine learning techniques (Hall *et al*., 2009).

The Weka workbench offers easy access to these capabilities through graphical stoner interfaces and a diversity of visualization tools and algorithms for data analysis and prophetic modeling. It's software that's open source. It can run on virtually any current computing platform because it's completely enforced in the java programming language, making it movable and platform-neutral. Data pre-processing, clustering, bracket, association, visualization and point selection are just a many of the common data mining conditioning that Weka can perform. Weka's graphical terrain is launched using the Weka GUI namer, which includes six buttons simple cli, explorer, experimenter, knowledge flow, ARFF viewer and log (Hall *et al*., 2009; Frank *et al*., 2004; Aher and Lobo, 2011).

Once the data is collected, the Weka visualization tool is employed for data analysis. Weka provides a rich set of data mining and machine learning algorithms, allowing for in-depth analysis of the collected labor data. Various techniques such as clustering, classification and visualization are utilized to gain insights into the characteristics of casual labor in Libya. The visualization capabilities of Weka enable the representation of complex labor data in a comprehensible manner, facilitating decision-making processes.

The system will be evaluated through real-world data from various sources, assessing the effectiveness of data collection and the insights gained through data analysis. In conclusion, the proposed system for collecting and analyzing casual labor data in Libya based on distributed systems and Weka visualization offers a novel approach to addressing the challenges of decentralized labor markets.

System Architecture

Casual workers in Libya are a very big problem due to a large number of irregular migrants from the neighboring countries of Libya and since Libya is a transit country for irregular migrants, especially immigrants to the countries of the European Union, in order to work and earn a living and provide a decent life for their families, for these reasons, employment expatriates work for immigration to European

Union countries (El-Fadil, 2009). The most characteristic characteristics of casual employment are the following:

1. The level of education is often low
2. Most casual workers work in professions that do not require great educational qualifications
3. Most of the casual laborers work as daily workers
4. Casual workers often have low wages

The problem of casual employment is that it affects economic life in the Libyan state, so the initiative must be taken to collect data on this group of people, study it, organize it and limit its spread. This research presents a proposal to study and analyze casual employment data using distributed systems with the Ray framework.

After getting acquainted with the problem of casual labor in Libya, we must briefly get acquainted with the Ray framework.

Ray is an open-source project of a language called Python for parallel programming and distributed systems. It is one of the modern systems and frameworks that helps to prepare software based on the logic of working parallelism and distributed systems, for which parallel and distributed processing are integral parts of modern applications (Hall *et al*., 2009; Frank *et al*., 2004; Aher and Lobo, 2011; Ray *et al*., 2023; Ditscherlein *et al*., 2022; Kashani *et al*., 2021).

This research needs to use multiple cores or multiple machines to accelerate applications or run them at scale. The software and network application infrastructure and query and query-response systems are not single-threaded programs running on someone's laptop but rather a set of services that communicate and interact with one another. distributed and running in parallel applications (Ditscherlein *et al*., 2022; Kashani *et al*., 2021). As depicted in Fig. 1.

In this project, the Ray framework will be used and defined locally within the computer because the Ray framework is a virtual environment containing 4 cores, each core representing a remote computer.

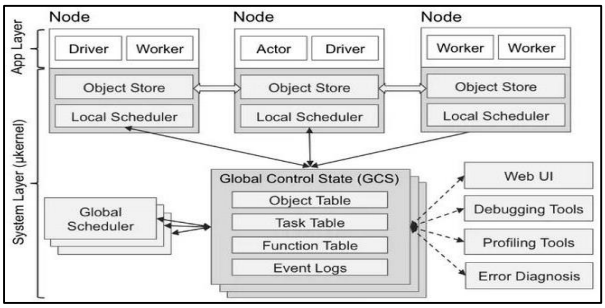

Fig. 1: Ray framework architecture

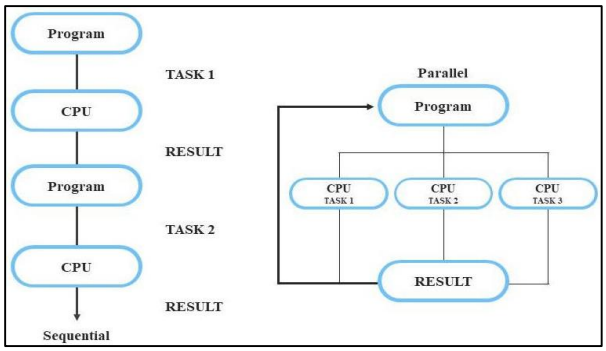

Fig. 2: Working parallel and sequential in the system

The idea of the project is mainly to use distributed systems, as shown in Fig. 2. The main idea of this project is to propose a system for collecting and analyzing casual labor data in Libya based on distributed systems. The importance of using distributed systems is the speed at which the study is performed while taking advantage of the power of data processing in parallel with the Ray framework in the Python language. Figure 2 illustrates the idea of working in parallel compared to sequential work in the system.

In traditional systems based on the central milling unit, operations are carried out sequentially, which is suitable for simple jobs. But for complex systems, which require a lot of effort and simultaneous data entry operations, the use of distributed systems is the solution to this problem.

Figure 2 parallel calculations are faster than the method of sequential calculations and the difference here is that working in parallel saves time and effort on the CPU, while working in a row consumes more resources in the system compared to working in parallel.

Next: Explanation of the casual labor program in Libya in Python using distributed systems.

GUI Implements the Project in Python

In the casual labor program in Libya, which was programmed in the Python language, at the start of running the program, a graphical interface appears, offering you two options to calculate the time for executing data processing, either in a serial way or in a distributed way. The program performs these operations in the background (Fig. 3). When a sequential command is implemented, the output of the execution appears in the background through the console, as shown in Fig. 4. When you press the button run distributed, which means in parallel, the output of the execution appears in the background through the console. The result is presented in Fig. 5.

After execution on the distributed system, execution in the background, the execution time appears and it is noted here that the execution time with distributed running is less than execution on sequential. The final result after implementation is presented in Fig. 6.

Fig. 3: The implementation of the proposed system of collecting and analyzing data using the Python language

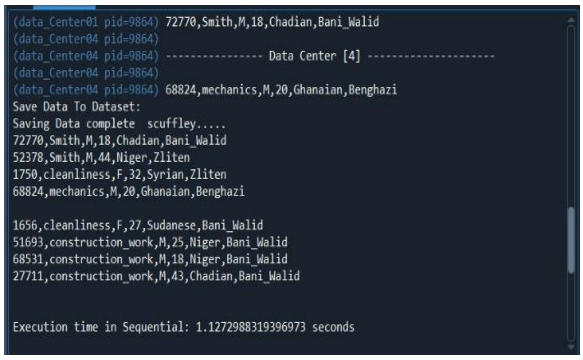

Fig. 4: Run in a sequential state

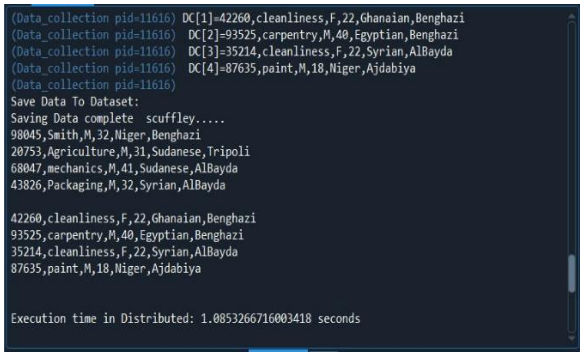

Fig. 5: Run in a distributed state

Fig. 6: Final results after running the GUI in the project

After the program is executed, the virtual data centers, which collect casual labor data, create a data set that is stored within the program during execution so that we can analyze this data using Weka programs.

Results and Discussion

Evaluation

After implementing the system and testing it, it turned out that there is an advantage: The implementation in the distributed system takes less time than the implementation in the sequential system, as shown in Fig. 7. Table 1 shows a sample of project implementation, depending on the execution time.

Each time the program is executed, the collected data is added to a file so that it can be analyzed later by Weka (Fig. 8). Next: Analysis of data collected by the casual workers analysis program in Libya.

Weka Evaluation

Weka is a powerful data mining application that can help you better understand the data you acquire (Hall *et al.* 2009) (Fig. 9) depicts the GUI chooser. The application features powerful data analysis tools that can be used to extract information and develop new machine-learning schemes. Overall, the app is an excellent data analysis tool. Figure 10 depicts Weka explorer within six attributes distances.

Table 1: Implementation, depending on the execution time main headings

ID	Sequential	Distributed	Difference
1	1.117308130	1.102317048	0.014991082
2	1.190263747	1.129300117	0.060963630
3	1.178915596	1.108314110	0.070601486
4	1.433320618	1.123306512	0.310014105
5	1.107308140	1.100564000	0.006744140
6	1.190263748	1.129300117	0.060963631
7	1.189155960	1.122460000	0.066695960
8	1.083332062	1.033065128	0.050266933

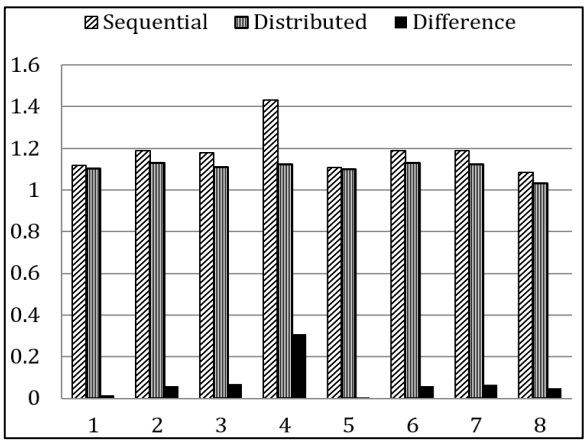

Fig. 7: Comparison of execution time in the program between sequential and distributed

ID	JOB	GENDER	AGE	NATIONALITY	CITY
1	Electricity_technician	M	41	Niger	Tripoli
2	blacksmith_shop	M	30	Niger	Tripoli
3	Electricity_technician	M	44	Niger	Tripoli
4	Packaging	M	23	Niger	Tripoli
5	mechanics	M	39	Niger	Tripoli
6	carpentry	M	37	Niger	Tripoli
7	blacksmith_shop	M	28	Niger	Tripoli
8	mechanics	M	20	Niger	Tripoli
9	Packaging	M	28	Niger	Tripoli
10	Packaging	M	24	Niger	Tripoli
11	daily_worker	M	23	Niger	Tripoli
12	cleanliness	F	26	Niger	Tripoli
13	daily_worker	M	30	Niger	Tripoli
14	Agriculture	M	35	Niger	Tripoli
15	mechanics	M	26	Niger	Tripoli
16	paint	M	35	Niger	Tripoli
17	cleanliness	F	21	Niger	Tripoli
18	daily_worker	M	44	Niger	Bani_Walid
19	construction_work	M	32	Niger	Bani_Walid
20	daily_worker	M	43	Niger	Bani_Walid
21	Electricity_technician	M	43	Niger	Bani_Walid
22	construction_work	M	33	Niger	Bani_Walid
23	daily_worker	M	41	Niger	Bani_Walid
24	Smith	M	28	Niger	Bani_Walid

Fig. 8: Data set provided from the data collection

Fig. 9: The GUI chooser

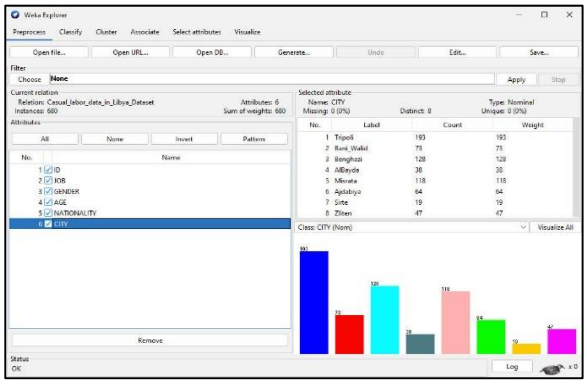

Fig. 10: Weka explorer within six attributes distances

Scalability and Reliability

After implementing a proposed system of collecting and analyzing casual labor data in Libya based on distributed systems and Weka visualization and evaluating the results, it became clear that I meant the advantages of working on parallel (distributed) systems using the Ray frame and it has been shown through practical experience that multi-core systems make optimal use of computer resources. Among the most important of these sources is the main memory. Parallelism gives the advantage of speed in completing the study and this was clearly shown in the results of the distributed system calculations compared to the sequential system. This provides better scalability than traditional CPU systems. Cloud computing can be used to speed up work completion, which greatly helps in making calculations and quickly evaluating the system. This depends on the cost of the project and the budget allocated for expanding the system.

Trade-off and Limitations

In this project, the project calculations are implemented based on the distributed systems, for the proposed system of collecting and analyzing casual labor data in Libya based on distributed systems and considering the use of a local version of the RAE framework, where the RAE framework provides a maximum of 4 cores to work on computers. Therefore, no more than 4 cores have been worked out to implement the system and the execution time calculations showed the advantage of using the distributed evaluation method compared to the successive calculations. Working on the principles of parallelism and computation in distributed systems remains one of the goals achieved in this project.

Simple K-mean (Fig. 11) and Weka visualization that analyzed the jobs occupied by casual labor in eight cities in Libya, namely, Tripoli, Bani Walid, Benghazi, Al-Bayada, Misrata, Ajdabiya, Sirt and Zliten. The results indicated that the highest casual labor job is auto electrician and the lowest is residential electrician. In addition, simple K-mean and Weka visualization that represented casual labor in different cities and their nationality showed that the highest casual nationality is Sudanese and the lowest are from Niger. Canopy and Weka visualization that represented casual labor in deferent cities verses jobs indicated that the highest casual jobs are auto electrician and the lowest are residential electrician. Also, hierarchical clustered and Weka visualization that analyzed casual labor city and their jobs showed that the highest casual jobs are auto electrician and the lowest are residential electrician.

The graphical interface offered by the program is a very intriguing feature. It is not as complex as ArcGIS and quantum GIS, but it still lets the user execute visual searches and see the data graphically via this interface. Weka is an open-source machine learning technique that can be called from pre-existing Java code or applied directly to a dataset for data mining jobs. Tools for pre-processing data, regression, classification, clustering, visualization and association rules are all included in the system. For creating novel machine learning schemes, it works well. With the use of filters, attribute selection classes and meta-classifiers, Weka users can carry out attribute selection. Using transaction ID as the primary key, the program lets users access transactional data kept in databases. By defining and adjusting the colors in the interface, teams may also alter the plots' background color (Engel *et al*., 2014; Hall *et al.*, 2009).

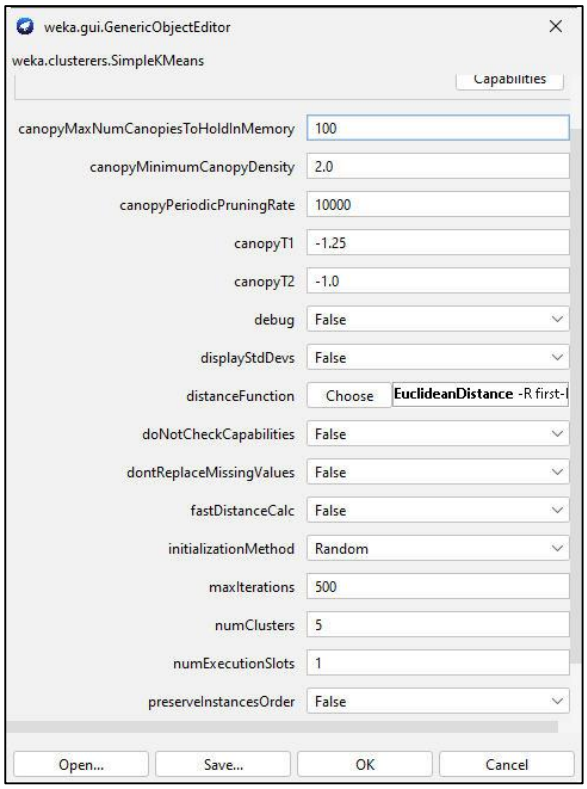

Fig. 11: Simple K-mean

Conclusion

After completing, the implementation of the proposed system of collecting and analyzing casual labor data in Libya based on distributed systems and Weka visualization. Four cores were used as many as the Ray frame allows, in the local version, in the PC, where each core represents a remote independent machine, running in parallel with the others. The importance of using computer resources, which is the advantage of memory management, increased performance and speed of work completion, was taken advantage of with the application of the principle of distributed systems that work in parallel using the Python language. Simple K-means clustering (casual labor city and their jobs) the highest casual labor jobs are represented by electrical (95%) as well as the lowest casual labor job which is the residential electrician (0%). In the same way, the simple K-means clustering (casual labor city and their nationality) represented that the highest casual nationality is Sudanese (98%), however, the lowest casual nationality is Niger (22%). In the same way, by canopy clustering (casual labor city and their jobs) which represented the highest number of casual jobs are swath electric (88%) and the lowest number of casual jobs: Residential electrician (0%). In addition, the hierarchical cluster represented the casual labor city and their jobs which the highest casual jobs are working in auto electric (97%) and the lowest casual jobs are working in electrical installation of residential buildings (7%).

Future Work

In the future, this research can develop a system for the proposed system of collecting and analyzing casual labor data in Libya based on distributed systems-based data warehousing and it is also possible to work on analyzing this data using deep learning systems and integrating this data with the economic movement within the city to obtain ratings, of high value, of the movement of transportation within the city.

Acknowledgment

Thank you to the publisher for their support in the publication of this research article. We are grateful for the resources and platform provided by the publisher, which have enabled us to share our findings with a wider audience. We appreciate the effort.

Funding Information

The authors would to disclose that the work presented here does not receive any significant financial support.

Author's Contributions

Magda Shuayb: Performed data analysis of the research. Also, participated in written the manuscript.

Ashraf Fadiel and Taher Abu-Lebdeh: Provided the research topic and guided the research development, research planed and data analysis. Also, participated in written the manuscript.

Ethics

The authors would to disclose that Dr. Taher Abu-Lebdeh and Ashraf Fadiel (Co-authors) are members of the editorial board for the American journal of engineering and applied sciences.

References

Aher, S. B., & Lobo, L. M. R. J. (2011). Data mining in educational system using Weka. *In International Conference on Emerging Technology Trends (ICETT), 3*, 20-25.

Bakthavatchalam, K., Karthik, B., Thiruvengadam, V., Muthal, S., Jose, D., Kotecha, K., & Varadarajan, V. (2022). IoT framework for measurement and precision agriculture: Predicting the crop using machine learning algorithms. *Technologies*, *10*(1), 13. https://doi.org/10.3390/technologies10010013

Bourechak, A., Zedadra, O., Kouahla, M. N., Guerrieri, A., Seridi, H., & Fortino, G. (2023). At the Confluence of artificial intelligence and edge computing in IoT-based applications: A review and new perspectives. *Sensors*, *23*(3), 1639. https://doi.org/10.3390/s23031639

Breza, E., Kaur, S., & Shamdasani, Y. (2021). Labor rationing. *American Economic Review*, *111*(10), 3184-3224. https://doi.org/10.1257/aer.20201385

Desiere, S., & Costa, V. (2019). Employment data in household surveys: Taking stock, looking ahead. *Looking Ahead. World Bank Policy Research Working Paper*, (8882). https://doi.org/10.1596/1813-9450-8882

Ditscherlein, R., Leißner, T., & Peuker, U. A. (2022). Preparation strategy for statistically significant micrometer-sized particle systems suitable for correlative 3D imaging workflows on the example of X-ray microtomography. Powder Technology, 395, 235-242. https://doi.org/10.1016/j.powtec.2021.09.038

El-Fadil, E. H. M. (2009). Higher education outputs and local labour-market requirements in Libya. *The University of Liverpool*, (United Kingdom). https://www.proquest.com/openview/a60a1fef2d72ad23aaa99ce6310ec062/1?pq-origsite=gscholar&cbl=2026366&diss=y

Engel, T. A., Charao, A. S., Kirsch-Pinheiro, M., & Steffenel, L. A. (2014). Performance improvement of data mining in Weka through GPU acceleration. *Procedia Computer Science*, *32*, 93-100. https://doi.org/10.1016/j.procs.2014.05.402

Frank, E., Hall, M., Trigg, L., Holmes, G., & Witten, I. H. (2004). Data mining in bioinformatics using Weka. *Bioinformatics*, *20*(15), 2479-2481. https://doi.org/10.1093/bioinformatics/bth261

Gary, K. (2019). The distinct seasonality of early modern casual labor and the short durations of individual working years: Sweden 1500-1800. *Lund Papers in Economic History*, *189*. https://portal.research.lu.se/en/publications/the-distinct-seasonality-of-early-modern-casual-labor-and-the-sho

Gomes, H. M., Read, J., Bifet, A., Barddal, J. P., & Gama, J. (2019). Machine learning for streaming data: State of the art, challenges and opportunities. *ACM SIGKDD Explorations Newsletter*, *21*(2), 6-22. https://doi.org/10.1145/3373464.3373470

Heikkonen, M. (2021). Automated machine learning: Evaluating Auto ML frameworks (Master's thesis). https://trepo.tuni.fi/bitstream/handle/10024/134846/HeikkonenMiro.pdf?sequence=2

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The Weka data mining software: An update. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10-18. https://doi.org/10.1145/1656274.1656278

Hamed, A., Sobhy, A., & Nassar, H. (2021). Distributed approach for computing rough set approximations of big incomplete information systems. Information Sciences, 547, 427-449. https://doi.org/10.1016/j.ins.2020.08.049

Kashani, M. H., Madanipour, M., Nikravan, M., Asghari, P., & Mahdipour, E. (2021). A systematic review of IoT in healthcare: Applications, techniques and trends. *Journal of Network and Computer Applications*, *192*, 103164. https://doi.org/10.1016/j.jnca.2021.103164

Manalu, J. P., Sukmawati, A., & Cahyadi, E. R. (2023). Strategy to increase participation of palm oil enterprises in registering casual laborers to labor social security program. *Jurnal Manajemen and Agribisnis*, *20*(1), 33-33. https://doi.org/10.17358/jma.20.1.33

Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., López García, Á., Heredia, I., ... & Hluchý, L. (2019). Machine learning and deep learning frameworks and libraries for large-scale data mining: A survey. *Artificial Intelligence Review*, *52*, 77-124. https://doi.org/10.1007/s10462-018-09679-z

Rado, O. A. (2019). Contributions to evaluation of machine learning models. *Applicability Domain of Classification Models* (doctoral dissertation, University of Bradford). https://bradscholars.brad.ac.uk/handle/10454/18447

Ray, S. S., Verma, R. K., Singh, A., Ganesapillai, M., & Kwon, Y. N. (2023). A holistic review on how artificial intelligence has redefined water treatment and seawater desalination processes. Desalination, 546, 116221. https://doi.org/10.1016/j.desal.2022.116221

Spaargaren, R. (2020). Engagement of casual labor to hotel arts' standards and culture. https://www.recercat.cat/handle/2072/373878