ISSN 2522-9400 **European Modern Studies Journal** Vol 8 No 6
available at www.journal-ems.com

DOI: 10.59573/emsj.8(6).2024.21

The Role of Machine Learning in Supporting the Decision in Power Sector-COMESA's Region-Sudan

Dr. Nessreen Abdelfatah Ali Abdoun[1], Eng. Khaled Mafouz[2]
[1]College of Engineering and Information Technology
Ajman University, Ajman, UAE
ORCID: 0000-0002-2991-5111
[2]College of Engineering and Information Technology
Ajman University, Ajman, UAE
ORCID: 0000-0003-1351-9007

Abstract. According to the huge amount of data generated by the power sector around the world and the challenges facing power sectors in Africa, this work provides a review of the existing situation in the power sector in Sudan as part of COMESA's region. Data mining and analytics have played an important role in knowledge discovery and decision-making in the process of the proposed reform and recovery strategy for the power sector in Sudan using ML techniques for information extraction, data pattern recognition that can handle the increase of the data. This work also presents the advantages of the Weka tool in making decisions regarding the ideal power generation projects according to the type of fuel and technologies and the installed capacity. It also gives a clear idea about the suitable business model for the sector to perform the required assignment that helps to achieve the purpose of the actions that should be implemented to support the power sector in Sudan.

Keywords: Machine Learning, Power sector, Weka software, Data Mining, Business Model

Introduction

The provision of reliable and affordable energy services is important for any country to address many of the development programs and challenges to achieve the United Nations Sustainable Development Goals and promote economic development. In Africa, despite the diversity of energy resources (many papers and reports detail the estimated resource potential for renewable energy, especially solar and wind), there are many barriers to accessing electricity due to lack of financial challenges, political situations, and lack of energy policy and regulatory frameworks. Electricity demand will continue to grow due to increasing development needs in the COMESA region (Lipimile, 2012; Common Market for Eastern and Southern Africa et al., 1998).

The study and analysis of historical data for the electricity sector in Sudan over the last decade (2008-2018) showed that electricity consumption has been steadily increasing, with the average peak load increasing by 13.5% and the number of consumers increasing from 1,186,030 consumers in 2008 to 2,604,858 consumers in 2018, an increase of 120%. This growth is a major challenge in providing the necessary resources for terminal expansion (generation, transmission, and distribution networks) to meet the rapid increase in demand (S. E. H. C. Ltd., 2021). International donors (the World Bank) have developed various strategies to reform and rehabilitate the power sector in Sudan through a diagnostic review conducted by the WB team. The team concluded that the sector is operating efficiently compared to regional standards, but the most pressing issue facing the sector is financial sustainability, which, along with four other challenges, can lead to power crises. The biggest challenge is the provision of accurate and verified data (S. E. H. C. Ltd., 2021).

In this work, the requirements of data mining in the Sudanese power sector were presented because Big Data is the term that describes the collection of large and complex data sets. Big Data brings several problems such as collection, storage, and sharing. The electricity sector in Sudan has a huge amount of data, but there is a gap in using appropriate knowledge extraction tools to benefit from it, and there are various algorithms in data mining that can help in decision making. In this paper, WEKA 3.8.6 has been introduced as a tool that can perform many data mining tasks such as pre-processing data, classification, and improving knowledge discovery. WEKA is an open-source Machine Learning Data Mining, Java-based software developed by the University of Waikato from New Zeeland which is free software for non-commercial purposes. Furthermore, Weka supports large model ratings and metrics and is more oriented towards classification and regression problems (Robu & Stoicu-Tivadar, 2010; Jovic, Brkic, & Bogunovic, 2014; Sharma, Bajpai, & Litoriya, 2012).

Literature Review

Data mining is very important when it comes to the explosive growth of data from terabytes to petabytes. The major problem in the power sector was the availability of data and many companies invested in building data warehouses that contain millions of records and attribute to support and encourage investment but still, they are not getting the Return on Investment. In addition to that, no sufficient output will be produced with the lack of knowledge, staff, and appropriate tools and the main task of data mining is to explore a large amount of data from a different point of view, classify it and summarize it (Singhal & Jena, 2013; Kulkarni & Kulkarni, 2016).

Many research has been conducted to enhance decision-making according to data mining and using different tools and techniques. One of the researchers' outputs in this field is that the difficulty of deciding on control action depends on the region of variables. Data mining methods have the best performance in the region of little decision difficulty and have the poorest performance in the region of most decision difficulty (Deepak Kulkarni, Yao, & Banavar, 2013). In every organization, the production of knowledge comes from the process of effective information especially in big and demanded businesses such as industries. The research that was conducted from the field of a Greek railway company has applied machine learning techniques to create strategic decision support and draw up a risk and control plan for trains using Weka software as a Machine Learning open-source software to the obsolete procedures of maintenance of the rolling stock of the company “Figure 1”. Data were recorded and analyzed using the J48 and M5P algorithms from the Weka software and this method is capable of the management’s performance of the trains to obtain accurate information that helps in the passengers’ safety (Kalathas & Papoutsidakis, 2021).

The process industry has gained lots of developments in the past several decades and sustainability has received great efforts in energy efficiency research, environmental pollution, and monitoring and treatment programs for the sewage. Machine learning can efficiently describe dynamic relationships among process data are particularly useful to real-time process control and optimization in operating performance, energy consumption, maintenance and emissions and the sustainability of the process can be predicted through observing some measure process. Therefore, the sustainability of the process when it comes to the variables and the prediction of components based on the data collected from sensors will help financially (Ge et al., 2017; Dogan & Birant, 2021).

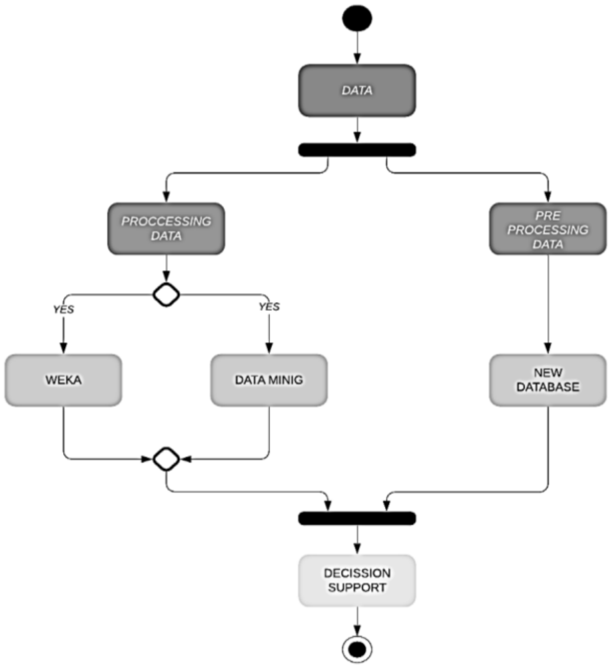

Figure 1. Flow chart for the decision-making process

In the power sector and because of the huge data, this sector needs to carry out the work safely and effectively. This is why some of the researchers recommended the design of a Knowledge Management System (KM). The proposed solution from the researcher was making a Hybrid Artificial Intelligent System (HAIS) that supports a Knowledge Management system by applying two bio-inspired techniques Artificial Neural Networks (ANNs) and Genetic Algorithms (GAs) in an innovative way “Figure 2” (SÃ¡ iz-BÃ¡ rcena et al., 2015).

The losses of electric power are considered an important topic and many engineers and scientists have been tried to find solutions that can help in reducing the technical and non-technical losses. From the research conducted in a Brazilian power company, data mining techniques are used to detect the non-technical losses in power distribution systems using Weka software to compare and classify various techniques through intelligent algorithms. The proposed method was validated on a set of labeled data from the company profile and the results proved that a subset of features can increase the accuracy rate through WEKA software (S. E. H. C. Ltd., 2021; Sankari & Rajesh, 2015).

The revolution of smart and green cities has stimulated some authors to provide a comprehensive overview of the data analytics landscape on the integration of the Electric vehicle to green smart cities as a future roadmap for data analytics needs and solutions for green communication. The authors used Weka-based decision as one of the project's schemes using New York City (NYC) demand data to build a decision support engine for power system operators through J48 and M5 algorithms from the Weka platform (Li et al., 2017).

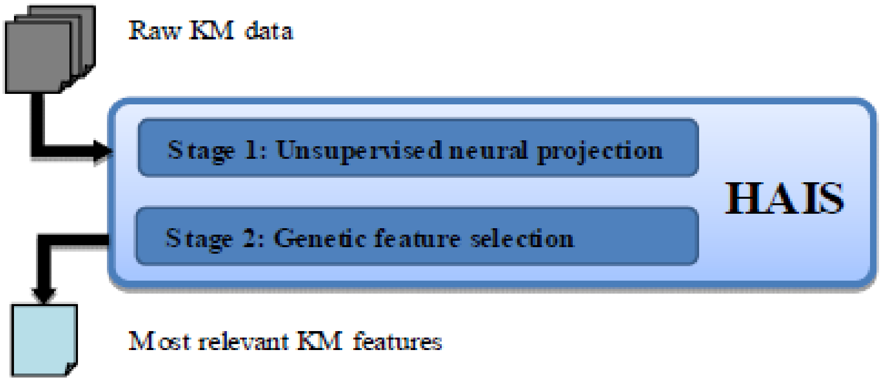

Figure 2. Proposed the two stages of HAIS solution

Methodology and Experimental Setup

The methodology used for this work is based on WEKA3.8.6 software to identify the ideal Power Generation Projects belonging to the process of Data Mining "Figure 3". This methodology contains very important stages; data collection, processing of collected data, data mining (classification), and interpretation "Figure 4". Data collection is gathering from the available information about the generation projects according to the generation expansion plan.

Data collection was gathered from the available information about the generation projects from different sources of information in the power sector then combined into the dataset. In the processing part, the collected data was incomplete and inconsistent to use in the Weka tool. WEKA tool accepts the data in Attributed Relation File Format (ARFF) and Comma Separated Values format (CSV). In this work, the CSV format is used as a data cleaning method.

In the classification phase, the DM algorithm is used for the classification of data for the best results. The interpretation phase is depending on the previous stages and it was analyzed to predict the generation projects through building a decision-making tree. Datasets used are created and saved as CSV files and then converted to ARFF format this file is an ASCII text file that describes a list of types with the values using a specific structure, because the Weka tool works effectively with ARFF files. ARFF files have two sections, the header information that contains the data information such as the relation's name and the attributes. In the experimental work, the classification techniques were applied to categorize the data for the most efficient use.

Figure 3. Weka3.8.6 Implementation tool

The parameters were analyzed according to the power technologies (as independent variables), fuel types (as control variables), and the installed capacity in Megawatts (as dependent variables) “Figure 5”.

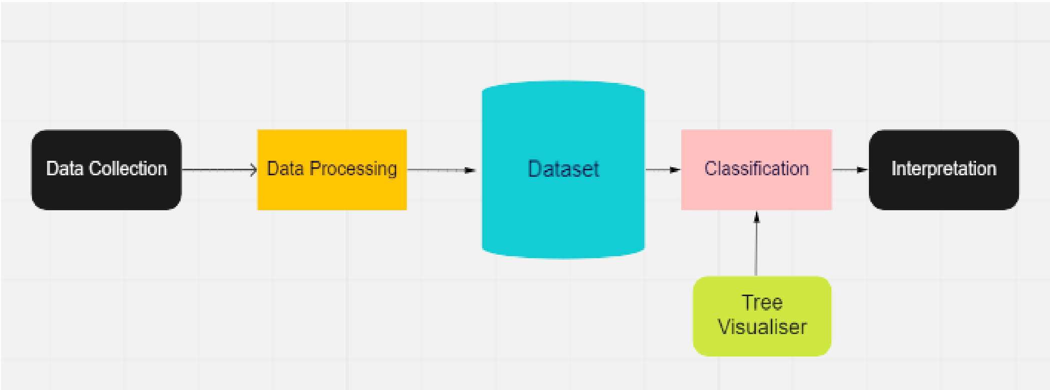

Figure 4. Proposed method to identify the ideal Power Generation Projects

Project	Technology	Fuel	ctive Commission	Quarter	Installed Capacity (MW)
Garri 3	OCGT	GO	2022	1st	561
Port Sudan Siemens	OCGT	GO	2023	1st	240
Darfur States Capital Cities	DE	HFO	2022	1st	150
Port Sudan Siemens	ST	HFO	2024	1st	175
Garri 3 Steam Siemens	CCGT	GO	2023	1st	240
Bagair 1	OCGT	GO	2023	1st	350
Bagair 1 Extension	OCGT	GO	2026	1st	175
Al Fulla II	CCGT	GO	2023	1st	450
Elshaheed	OCGT	GO	2024	1st	350
Bagair 2	CCGT	GO	2026	1st	225
Garri 4 Extension	ST	HFO	2026	1st	100
Kosti Extension	ST	HFO	2027	1st	500
Port Sudan Coal Red Sea	ST	HFO	2028	1st	600
Port Sudan Coal Red Sea 2	ST	HFO	2029	1st	300
Solar	Renewable	-	2021 – 2023	1st	610
Wind	Renewable	-	2023 – 2029	1st	281
Rosairis (RahadChennel)	Water	-	2027	1st	52
Kajbar Dam	Water	-	2028	1st	360
TOTAL					5,853

Figure 5. Generation Expansion Plan-Sudan

Results

In this work, the dataset is tested and analyzed with three Classification algorithms J48 pruned tree, REP Tree and Random Tree using a full training data set to select the accurate and efficient classifier (Table 1). After loading the dataset, the Weka tool has automatically presented the list of the attributes on the window with the displaying of the features without filtering “Figure 6”. By comparing the accuracy of the all classifiers “Figure 7”, “Figure 8”, Figure 9”, it has been investigated that the Random Tree technique has the best performance and correctly classified Instances (100%).

Table 1. The compression between the Classification Algorithms

Classification Algorithms	**J48 pruned tree**	**REP Tree**	**Random Tree**
Size of the tree	1	1	15
Time taken to build model	0 second	0.04 seconds	0.02 seconds
Correctly Classified Instances	83.3333 %	83.3333 %	100 %
Incorrectly Classified Instances	16.6667 %	16.6667 %	0 %

Kappa statistic	0	0	1
Mean absolute error	0.2778	0.2778	0
Root mean squared error	0.3727	0.3727	0
Relative absolute error	92.5926 %	92.5926 %	0 %
Root relative squared error	99.6024 %	99.6024 %	0 %
Total Number of Instances	18	18	18

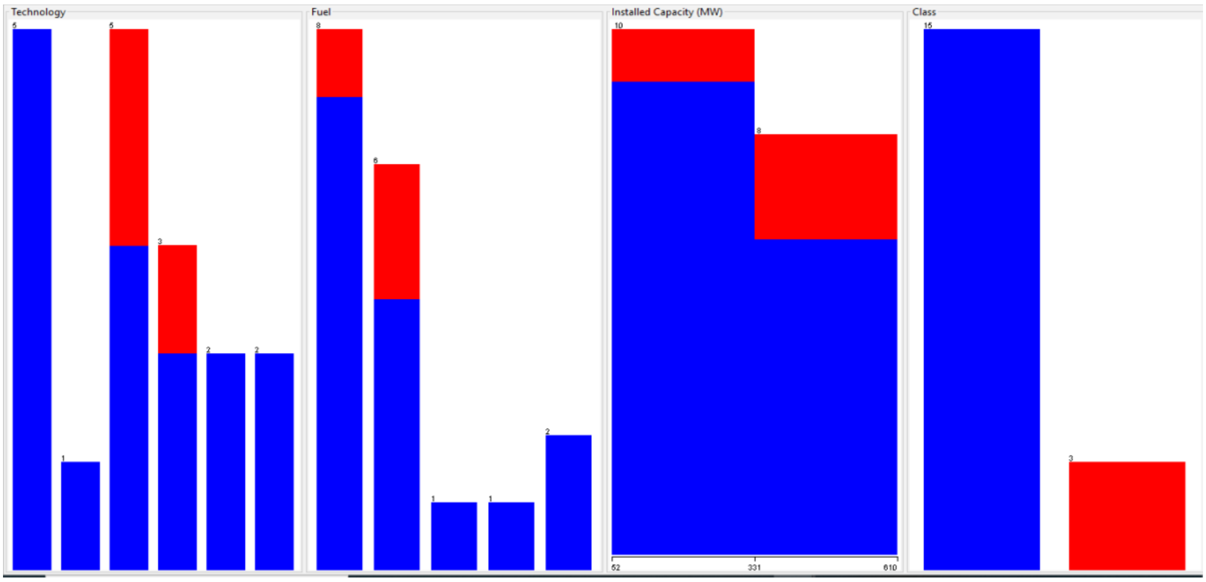

Figure 6. The visualization of the features

Random tree is built the decision using information gain without pruning with the presence of the estimated class probabilities according to the holdout set but by using J84 and RepTree, pruning is found using reduced error pruning.

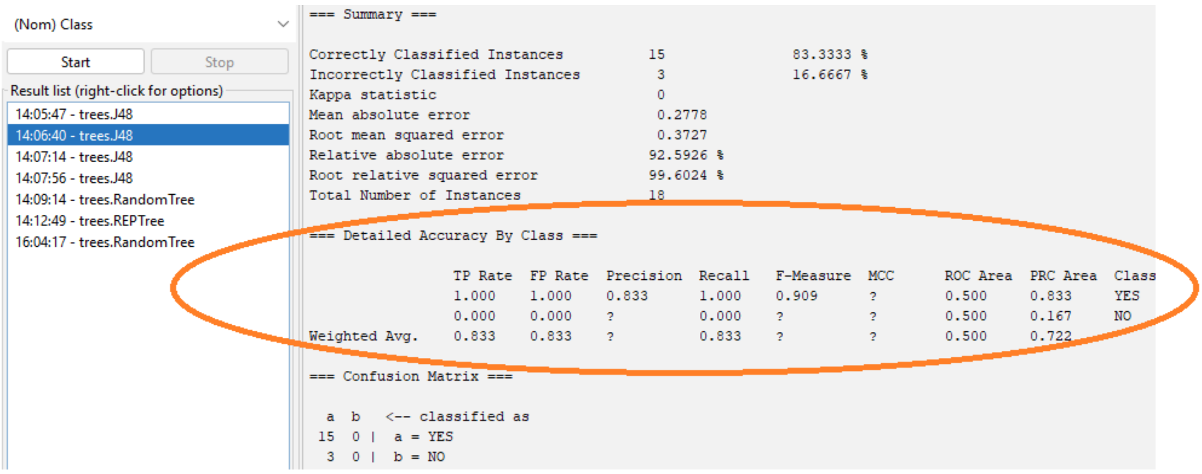

Figure 7. Data classification using J84

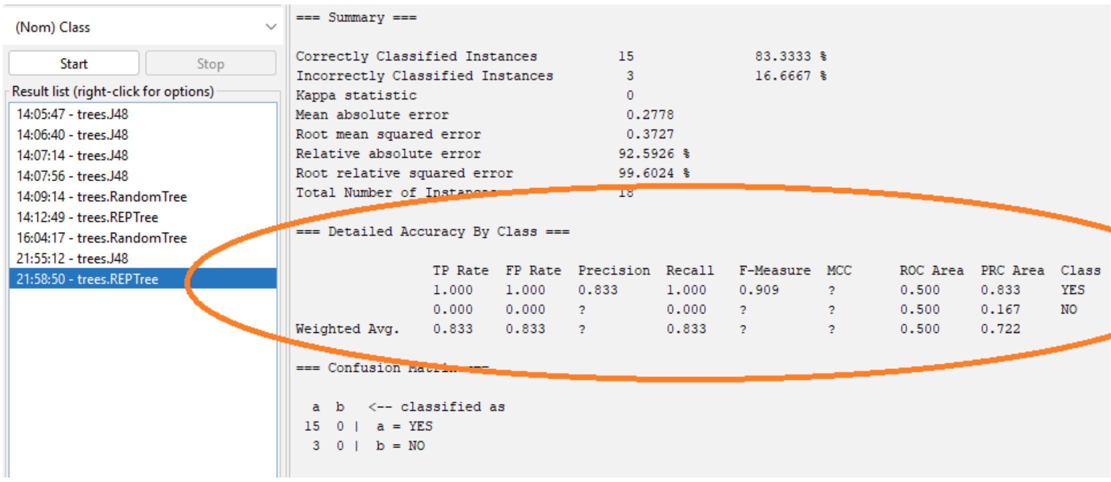

Figure 8. Data classification using Rep Tree

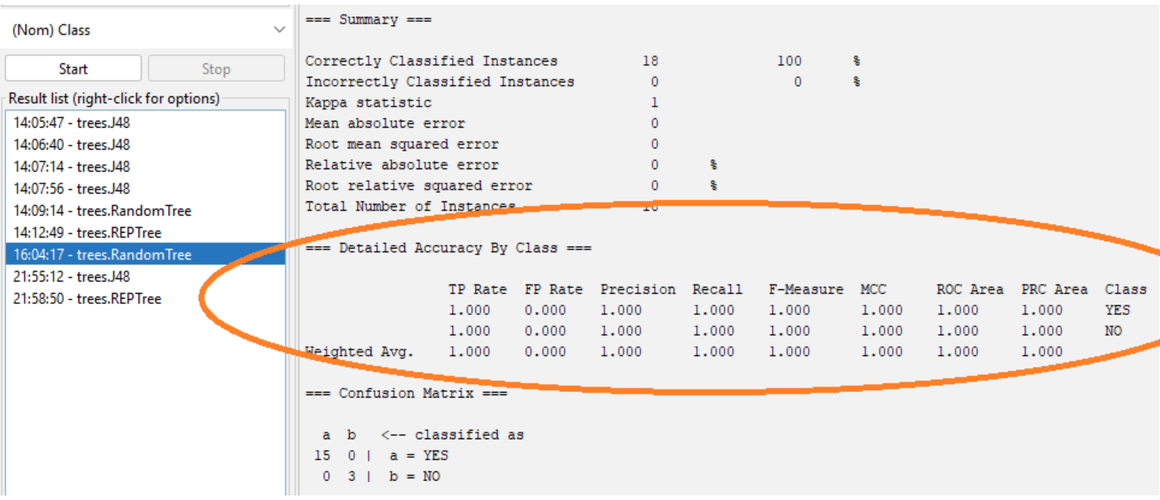

Figure 9. Data classification using Random Tree

In the visualization tree, the values on the lines that join nodes is representing the splitting criteria according to the values in the parent node feature and that gives a clear idea of the classification and regression problems. Based on the classification of the technologies in the generation projects, the technology with the answer NO answer was given another option according to the installed capacity. For example; If the technology with fuel ST is more than 237.5 the project should be taken but if its less the technology will be refused (Figure 10).

The business model for the power sector in Sudan was proposed based on the results of this work that showed the importance of data mining in taking decisions to provide the investors the knowledge about the competitive opportunities to invest in the power sector in Sudan the future expansion through the effective decision. This work has proposed the ideal power sector model by highlighting the regulatory framework with the database system (Figure 11).

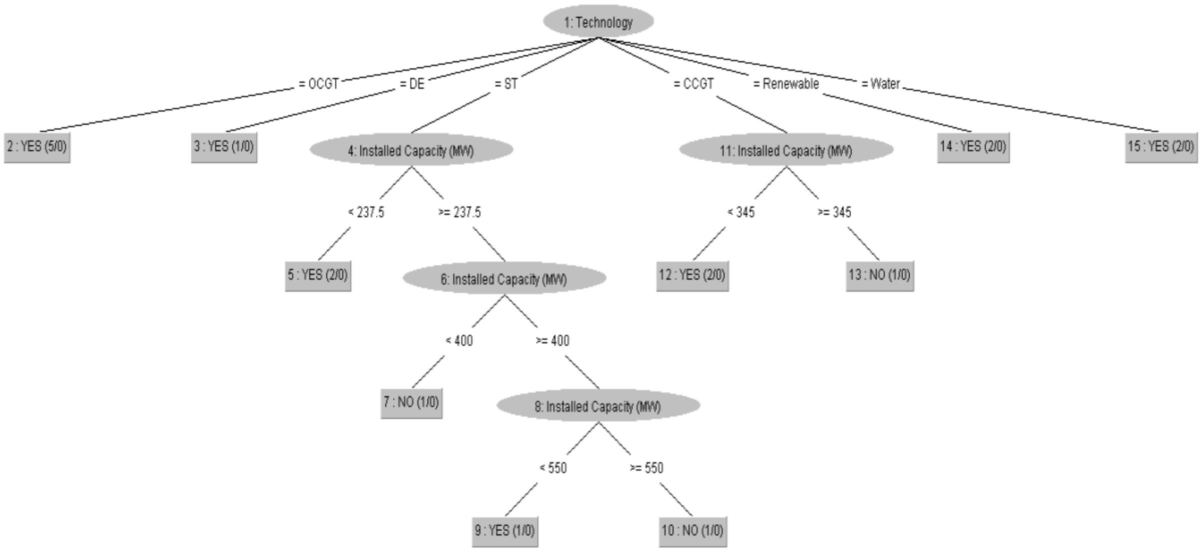

Figure 10. Visualize Tree-Random Tree

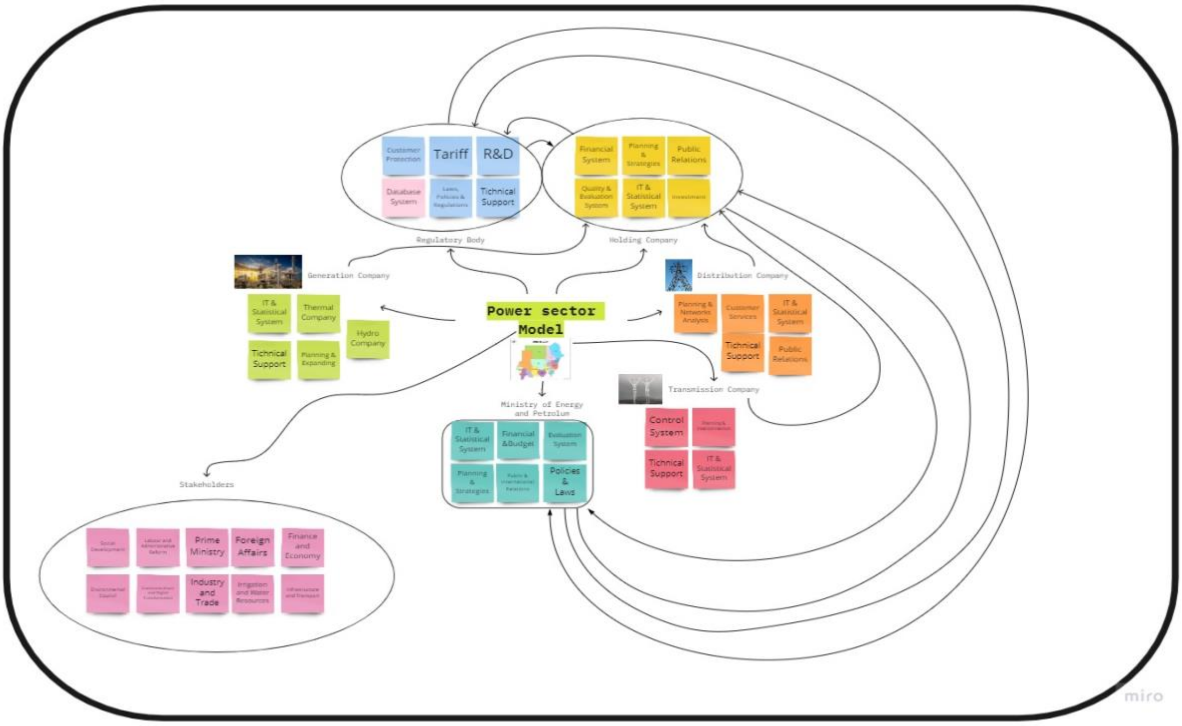

Figure 11. The Ideal Power sector model

Conclusion

This work can significantly support Sudan to increase the opportunities of generating electricity, providing an encouraging investment environment, and becoming a green country by reducing emissions of CO2 through Renewable energy projects. This article provides a comprehensive overview of the importance of data mining and the proposed business model. The future work can be building a business model for COMESA's countries that will enable data-driven investments and policy-making.

Acknowledgment

Common Market for Eastern and Southern Africa (COMESA), College of Engineering and Information Technology, Ajman University-UAE, High Class International Trade Limited, Sudan University of Science and Technology-Sudan and my father and my mother.

References

Common Market for Eastern and Southern Africa; United Nations. Economic Commission for Africa; United Nations. Economic Commission for Africa. Southern Africa SubRegional Development Centre (ECA/SA-SRDC) (1998-10). Progress report on priority programmes of COMESA. UN. ECA Intergovernmental Committee of Experts for Southern Africa (ICE) Meeting (5th : 1998, Oct. 05-08 : Lusaka, Zambia). Addis Ababa.

Deepak Kulkarni, K., Yao, W., & Banavar, S. (2013). *Data mining for understanding and improving decision-making affecting ground delay programs*. 2013 IEEE/AIAA 32nd Digital Avionics Systems Conference (DASC).

Dogan, A., & Birant, D. (2021). Machine learning and data mining in manufacturing. *Expert Systems with Applications*, *166*, 114060.

Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017). Data mining and analytics in the process industry: The role of machine learning. *IEEE Access*, *5*, 20590-20616.

Jovic, A., Brkic, K., & Bogunovic, N. (2014, May). An overview of free software tools for general data mining. In *2014 37th International convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 1112-1117). IEEE.

Kalathas, I., & Papoutsidakis, M. (2021). Predictive maintenance using machine learning and data mining: A pioneer method implemented to Greek railways. *Designs*, *5*(1), 5.

Kulkarni, G. E. & Kulkarni, B. R. (2016). WEKA powerful tool in data mining. *International Journal of Computer Applications*, *975*, 8887.

Li, B., Kisacikoglu, M. C., Liu, C., Singh, N., & Erol-Kantarci, M. (2017). Big data analytics for electric vehicle integration in green smart cities. *IEEE Communications Magazine*, *55*(11), 19-25.

Lipimile, G. K. (2012). The COMESA regional competition regulations, In J. Drexl, M. Bakhoum, E. M. Fox, M. S. Gal, & D. J. Gerber (Eds.), *Competition Policy and Regional Integration in Developing Countries*. Edward Elgar Publishing.

Robu, R., & Stoicu-Tivadar, V. (2010, May). Arff convertor tool for WEKA data mining software. In *2010 International Joint Conference on Computational Cybernetics and Technical Informatics* (pp. 247-251). IEEE.

S. E. H. C. Ltd. (2021). ELECTRICITY SECTOR PLAN, Implementation Strategy (2021-2035). Khartoum.

SÃ¡ iz-BÃ¡ rcena, L., Herrero, Á. L., Campo, M. Á. N. M. D., & Martínez, R. D. O. (2015). Easing knowledge management in the power sector by means of a neuro-genetic system. *International Journal of Bio-Inspired Computation*, *7*(3), 170-175.

Sankari, E., & Rajesh, R. (2015). Detection of non-technical loss in power utilities using data mining techniques. *International Journal for Innovative Research in Science & Technology*, *1*(9), 97-101.

Sharma, N., Bajpai, A., & Litoriya, M. R. (2012). Comparison the various clustering algorithms of weka tools. *Facilities*, *4*(7), 78-80.

Singhal, S., & Jena, M. (2013). A study on WEKA tool for data preprocessing, classification and clustering. *International Journal of Innovative technology and exploring engineering (IJItee)*, *2*(6), 250-253.