POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

ISSN: 1302-0900 (PRINT), ISSN: 2147-9429 (ONLINE)

URL: http://dergipark.org.tr/politeknik

Improvement of classification algorithms for energy saving in lost energy data of Libya electricity company using Weka model

Weka modeli kullanılarak Libya elektrik şirketinin kayıp enerji verilerinde enerji tassarufu için sınıflandırma algoritmalarının iyileştirilmesi

Yazar(lar) (Author(s))*: Ashraf Mohammed ABUSIDA [1], Secil KARATAY[2], Rezvan REZAEIZADEH[3], Aybaba HANCERLIOGULLARI[4*]*

ORCID[1]: 0000-0002-9735-5814

ORCID[2]: 0000-0002-1942-6728

ORCID[3]: 0000-0001-6219-6174

ORCID[4]: 0000-0001-7008-480X

To cite to this article: Abusida A.M., Karatay S., Rezaeizadeh R. and Hancerliogullari, A., "Improvement of classification algorithms for energy saving in lost energy data of Libya electricity company using Weka model", ***Journal of Polytechnic***, 26(4): 1697-1703, (2023).

Bu makaleye şu şekilde atıfta bulunabilirsiniz: Abusida A.M., Karatay S., Rezaeizadeh R. ve Hancerliogullari, A., "Improvement of classification algorithms for energy saving in lost energy data of Libya electricity company using Weka model", ***Politeknik Dergisi***, 26(4): 1697-1703, (2023).

Erişim linki (To link to this article): http://dergipark.org.tr/politeknik/archive

DOI: **10.2339/politeknik.1368126**

Improvement of Classification Algorithms for Energy Saving in Lost Energy Data of Libya Electricity Company Using Weka Model

Highlights

- *Comparison of the classification algorithms for the energy saving*
- *Weka model for the energy saving*
- *Optimization of the classification algorithms applied to SCADA database*

Graphical Abstract

The main goal of this study is to compare the performance of the classification algorithms applied to the Supervisory Control and Data Acquisition (SCADA) database in order to prevent the energy losses in General Electricity Company of Libya (GECOL).

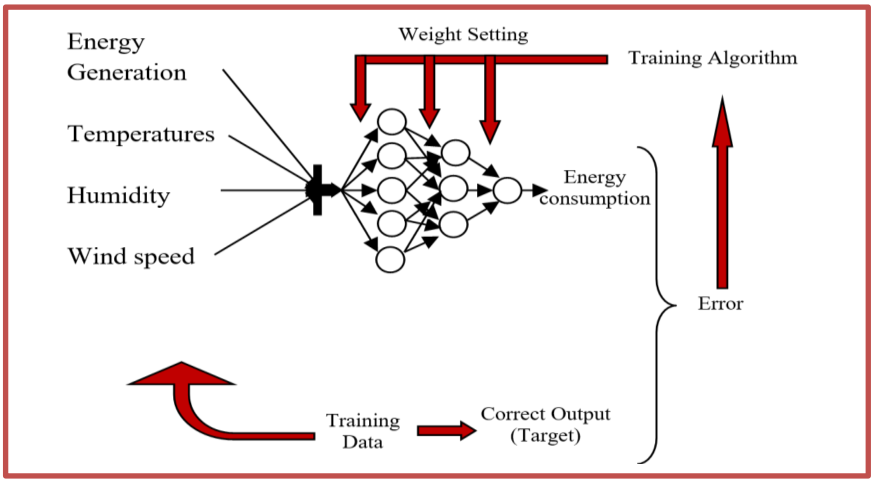

Figure. Architecture of MLP with backpropagation algorithm used in the study

Aim

The aim of to study is to prevent the energy losses in Libya electricity company and to improve the classification algorithms using SCADA data for this purpose.

Design & Methodology

The well-established data mining and classification software package known as the WEKA tool is used to minimize the losses.

Originality

With this study, six different parameters, including power production size, energy production time, energy production date, ambient temperature, humidity level and wind speed are used as the necessary data input for the algorithms forLibyan electricity company for the first time.

Findings

The Random Forest algorithm stands out as the best performing algorithm, achieving an impressive accuracy rate of 94.651% with a fast learning time of 9.56 seconds.

Conclusion

In cases where the energy consumption is high, it is possible to make accurate and rapid decisions regarding the determination and classification of time periods related to energy consumption.

Declaration of Ethical Standards

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Improvement of Classification Algorithms for Energy Saving in Lost Energy Data of Libya Electricity Company Using Weka Model

Araştırma Makalesi / Research Article

Ashraf Mohammed ABUSIDA[1], Secil KARATAY[2], Rezvan REZAEIZADEH[3], Aybaba HANCERLIOGULLARI[4*]

[1]Department of Materials Science and Engineering, Institute of Science, Kastamonu University, Kastamonu, Turkey
[2]Department of Electrical Electronics Engineering, Faculty of Engineering and Architecture, Kastamonu University, Kastamonu,Turkey
[3]Guilan University, Science Faculty, Rasht, Iran
[4]Department Physics, Faculty of Science, Kastamonu Unversity, Kastamonu, Turkey
(Geliş/Received : 28.02.2023 ; Kabul/Accepted : 26.04.2023 ; Erken Görünüm/Early View : 25.12.2023)

ABSTRACT

The main goal of this study is to compare the performance of the classification algorithms applied to the SCADA database of the Supervisory Control and Data Acquisition (SCADA) system of the General Electricity Company of Libya (GECOL). The company's annual energy and material losses have become seriously important to the Libyan government's research field. The well-established data mining and classification software package known as the WEKA tool is used to minimize these losses,. As necessary data input for algorithms; six different parameters are taken into consideration, namely power production size, energy production duration, energy production date, ambient temperature, humidity level and wind speed. This study is examined in detail for the first time in this article. In addition, considering the temperature variables, humidity, wind and other atmospheric effects of the environment, the energy losses of the company and the country are reduced to a minimum level. As a result, the company's annual electricity consumption is classified as low, medium or high consumption with the simulations. Thus, in cases where energy consumption is high, it is possible to make accurate and rapid decisions regarding the determination and classification of time periods related to energy consumption.

Keywords: Data mining, classification, SCADA, WEKA, GECOL.

Weka Modeli Kullanılarak Libya Elektrik Şirketinin Kayıp Enerji Verilerinde Enerji Tassarufu İçin Sınıflandırma Algoritmalarının İyileştirilmesi

ÖZ

Bu çalışmanın temel amacı, Libya Genel Elektrik Şirketi'nin (GECOL-General Electricity Company of Libya) Denetleyici Kontrol ve Veri Toplama (SCADA-Supervisory Control and Data Acquisition) sisteminin veri tabanına uygulanan sınıflandırma algoritmalarının performansını karşılaştırmaktır. Şirketin yıllık enerji ve malzeme kayıpları, Libya hükümetinin araştırma alanı açısından ciddi önem taşımaktadır. Bu kayıpları en aza indirgemek için WEKA aracı olarak bilinen köklü veri madenciliği ve sınıflandırma yazılım paketi kullanılmıştır. Algoritmalar için gerekli veri girişi olarak; güç üretim büyüklüğü, enerji üretim süresi, enerji üretim tarihi, ortam sıcaklığı, nem seviyesi ve rüzgâr hızı olmak üzere altı farklı parametre göz önünde tutulmuştur. Bu çalışma ilk kez detaylı olarak, bu makalede incelenmiştir. Bununla birlikte sıcaklık değişkenlerine göre ayrıca ortamın nem, rüzgâr ve diğer atmosferik parametreler de dikkate alınarak, şirketin ve ülkenin enerji kayıpları en az seviyeye indirgenmiştir. Sonuç olarak yapılan benzetimlerle, firmanın yıllık elektrik tüketimi düşük, orta veya yüksek tüketim olarak sınıflandırılmıştır. Böylelikle, enerji tüketiminin yüksek olduğu durumlarda, enerji tüketimine ilişkin zaman dilimlerinin belirlenmesi ve sınıflandırılması konusunda doğru ve hızlı kararlar alınmasına imkan sağlanmıştır.

Anahtar Kelimeler: Veri madenciliği, sınıflandırma, SCADA, WEKA, GECOL.

1. INTRODUCTION

Machine learning techniques have garnered substantial popularity in recent years and have emerged as a prominent subject of research and study, particularly within the domain of data mining. The data mining technique aims to reveal the information hidden in a huge of data after a series of specific operations, which were not accessible by traditional methods. The challenge at hand arises from the substantial reservoir of historical data stored within the data base of General Electricity Company of Libya (GECOL), which is integrated into the Enterprise Resource Planning (ERP) data set [1, 2].

**Sorumlu Yazar (Corresponding Author)*
e-posta : aybaba@kastamonu.edu.tr

SCADA (Supervisory Control and Data Acquisition) is an industrial computing system designed for real-time monitoring, control, and comprehensive event logging within the primary power plants, transmission and distribution infrastructure of electrical utilities, and the monitoring of substations, transformers, and other electrical assets within the company's infrastructure. In the context of this study, machine learning algorithms have been applied to analyze SCADA data, which consists of historical records continuously captured over twenty-four hours. These records encompass key parameters such as the total energy production from primary power stations, the overall network load, event timestamps, average temperature, humidity levels, weather conditions, and wind speed during these historical periods. Moreover, the employed technology aims to classify and predict the expected energy consumption levels, distinguishing between low, medium, and high demand, based on specific time periods, including peak hours, and considering various seasonal and weather-related factors. The primary objective is to uphold the stability of the electrical grid and mitigate production costs in response to demand fluctuations. Regrettably, the extensive and diverse data collections, stored in disparate structures, have rapidly proliferated to a point where they pose challenges for efficient utilization and management [3]. Furthermore, this approach can offer invaluable support to the company's decision-makers, enabling them to make timely and informed decisions. Consequently, it facilitates the company in upholding the quality of services it delivers to its customers [2, 4, 5]. Within the realm of machine learning algorithms, particular focus has been directed towards the application of classification algorithms to the SCADA dataset of the electricity company [6]. In contemporary times, applied information mining methodologies, as referenced in [7], have gained widespread adoption for the purpose of discovering novel and comprehensive datasets. The information mining process involves the extraction of discernible patterns from a given data source, resulting in the generation of multiple patterns and insights.

The General Electricity Company of Libya (GECOL) was established under the provisions of Act No. 17 of 1984 AD. This organization holds the responsibility for overseeing a comprehensive spectrum of operations, encompassing the construction, maintenance, and management of electrical utilities, power generation facilities, distribution substations, as well as the transmission infrastructure. Moreover, GECOL is tasked with the oversight of energy transmission lines and their distribution, power control centers, and the supervision of operations and maintenance for desalination stations across the entire nation [2, 8].

Libya is almost 100% electrified and it has one of the most robust power systems in Africa. Libyan national grid is divided into four operating areas with a total of 10 GW of generation installed capacity, with 15 power Plants containing 85 generating units of various sizes, technologies and ages. The Libyan power transmission network is distributed across the Country with a total length of about 16,000 km, with voltage levels of 400 kV, 220 kV and 132 kV. This transmission system is considered as a backbone to transfers the power to fed domestic, commercial agriculture and industrial Libyan power load demand of a total of about 8 GW [9]. In this study, a comprehensive optimization of the classification algorithms of machine learning is presented in order to be used the reducing energy losses in Libya. The WEKA model is used for the energy saving in the lost energy of GECOL. The Data Mining, the Material and Method and the Results are giving in Sections 2, 3 and 4, respectively.

2. DATA MINING

Data mining, defined as the practice of scrutinizing extensive preexisting databases with the objective of uncovering novel insights and knowledge, constitutes a pivotal process in the realm of pattern discovery from vast datasets. This procedure involves the application of intelligent methodologies to discern patterns within data. Data mining stands as an interdisciplinary subfield within the domain of computer science [1]. In a concise explanation, data mining refers to the extraction of meaningful and practical information from a larger collection of raw data. It encompasses the process of scrutinizing data patterns within extensive datasets, typically employing one or more software tools for analysis [2, 10]. Indeed, data mining is often referred to as Knowledge Discovery in Data (KDD). It entails the systematic process of gathering, searching, and analyzing extensive datasets within a database, with the aim of uncovering hidden patterns, relationships, or valuable knowledge.

2.1. Data Mining Techniques

Data mining is, at its core, the process of unearthing patterns and relationships within extensive datasets, achieved by leveraging algorithms and statistical analysis. In the realm of data mining projects, various major techniques have been developed and employed, encompassing association rules, classification, clustering, prediction, and pattern recognition. These techniques serve as essential tools for extracting valuable insights from data.

Indeed, association rule mining stands as a well-established data mining technique, particularly notable for uncovering meaningful association relationships among objects within a database. This technique primarily involves the generation of frequent item sets, which are sets of items that tend to occur together. In contemporary times, retailers have harnessed association rule mining to gain profound insights into customer purchasing behaviors. By analyzing these patterns, retailers can enhance their services, tailor their marketing strategies, and ultimately boost sales and company growth, all with the aim of better serving their customers. Also, companies and factories use association rule technique to search for a relationship between spare parts for machines and industrial and production equipment to take this into account when concluding deals for the purchase of spare parts, in order to speed up and improve the quality of services provided by these companies [4].

Clustering is a fundamental data mining technique characterized by the process of organizing data points into clusters or groups based on their intrinsic similarity or proximity to each other [11, 12]. This method allows for the identification of underlying structures or patterns within the data, making it valuable for various applications [13].

Prediction is a data mining technique that involves using a model to predict a numerical value for a new data point. In this technique, a model is trained using a dataset that includes a set of predictor variables and a numerical response variable [14]. The primary objective of a predictive model is to acquire an understanding of the relationship between predictor variables and the response variable. By doing so, the model aims to achieve the capability to make precise predictions for new, unseen data points [15]. For instance, a predictive model could be employed to forecast the price of a house, drawing on factors such as its location, size, and various other features as input variables.

Pattern recognition is a data mining technique that involves identifying patterns and trends in data. This can be used to identify fraudulent activity, predict customer behavior, or to identify trends in financial markets [16].

3. MATERIAL AND METHOD

A renowned data mining software package is the WEKA tool, short for "Waikato Environment for Knowledge Analysis." This widely used suite of machine learning software is Java-based and was developed at the University of Waikato, New Zealand. The origins of WEKA date back to 1993 when the University of Waikato initiated its development [4]. The system offers an extensive repertoire of potent machine learning algorithms tailored for data mining endeavors, encompassing some that are not readily available in commercial data mining systems. These encompass fundamental statistical capabilities and visualization tools, along with utilities for data pre-processing, classification, and clustering. Importantly, these resources are accessible through an intuitively designed graphical user interface, enhancing their user-friendliness and accessibility. Machine learning revolves around the development of computer systems with the capacity to enhance their performance within a specified domain through experiential learning. Both machine learning and data mining have emerged as progressively vital domains within the realms of engineering and computer science. Notably, these disciplines have been effectively applied to address a diverse array of challenges across the spectrum of science and engineering [4].

3.1. Data Set

The data set utilized for the experiments in this paper comprises historical records encompassing the total electrical energy production per hour, alongside corresponding date and time information, as well as concurrently recorded temperature, humidity, and wind speed data, all encapsulated within a defined time frame, as presented in Table 1.

As an example of the collected data employed in this study, Figure 1 provides a simplified representation of a random data sample, which includes the amount of energy consumed whether it is low consumption, medium consumption or high consumption, distributed over different times and under different climatic conditions.

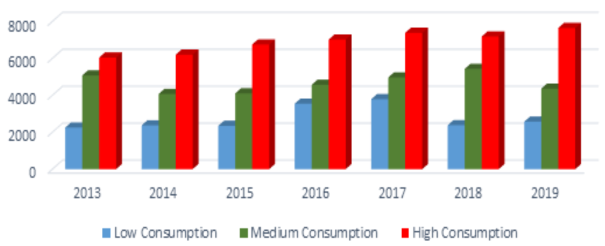

Figure 1. Random of data sample (law, medium and high consumption)

The date set consists of attributes related to amount of power generation, date, time and weather conditions such as temperatures, humidity and wind speed and level of power consumption whether low, medium or high consumption. Description of attributes and Datatype in the date set illustrate the input format of the data set, which is in ARFF (Attribute Relation File Format), a format commonly used as input for the WEKA tool.

Table 1. Properties of the SCADA data set

Variables	Unit	Description
Date	Date time	Date (Day-Month-Year)
Time	Time	Time (Hour: Minute: Second)
Energy Generation	Numeric	Energy Generation (Megawatts, MW)
Temperature	Numeric	Temperature (Degrees Celsius)
Humidity	Numeric	Humidity (%)
Wind Speed	Numeric	Wind Speed (Kilometers per Hour, km/h)
Energy Consumption	String	low, medium, high

3.2. Experimental Configuration

The tests in this study are conducted in two distinct modes:

- 10-Fold Cross-Validation: In this mode, the dataset is partitioned into ten subsets, and the model is trained and tested ten times, with each subset serving as the test data once while the remaining subsets aree used for training in each iteration.
- Evaluation on Full Training Data: In this mode, the model is trained using the entire data set, and its performance is evaluated based on the same data set. This mode represents a different approach where the model is assessed on its ability to generalize from the training data to unseen instances.

4. THE RESULTS AND THE DISCUSSION

WEKA is an efficient data mining tool with a vast capacity of analysis. It has inbuilt capability of analyzing Big Data using data mining techniques, It includes a collection of algorithms and tools for tasks such as: classification, clustering, regression and visualization. In this study, the analysis exclusively focuses on the classification technique. The classification approach encompasses seven distinct types of classifiers, namely Trees, Rules, Bayes, Functions, Lazy, Meta, and other. From this array of classifiers, four specific classifiers have been deliberately chosen for implementation and evaluation; Functions, Meta, Rules and Trees.

The WEKA application is chosen as the platform for implementing specific algorithms. The rationale behind this choice is to facilitate the comprehension of fundamental concepts and the practical application of these algorithms in real-time scenarios. WEKA proves valuable in grasping the fundamental principles of machine learning, offering diverse options for experimentation, and enabling the analysis of the generated results and outputs.

The WEKA program is employed, and a pre-processed and filtered data file that can be introduced into WEKA, as demonstrated in Figures 2 and 3.

The implementation of the classification algorithms is carried out according to the following steps. In the practical experiments carried out using WEKA Tools, twelve classification algorithms are applied to the data set and evaluation metrics are computed for each algorithm. During the implementation of classifiers on the real data set, two testing modes are taken into account:

- 10-Fold Cross-Validation: In this mode, the dataset is divided into ten subsets, and each classifier is evaluated using this cross-validation technique. The evaluation criteria includes the number of correctly classified instances and the time taken for each classifier.
- Full Training Set: In this mode, the classifiers are trained and tested using the entire data set as a single unit. The evaluation is again based on the correct classification of instances and the time taken.

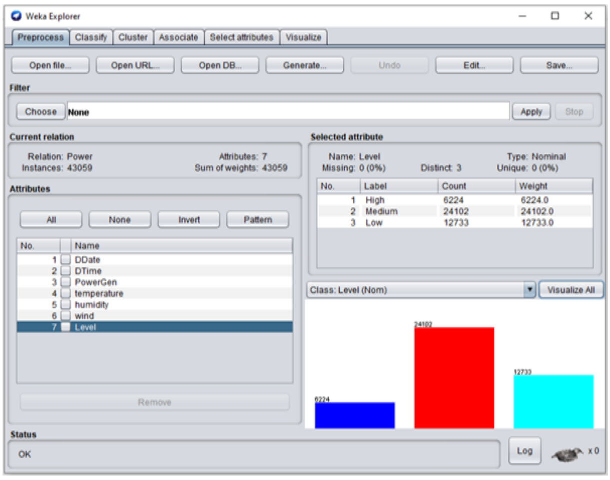

Figure 2. Load power arff file of random

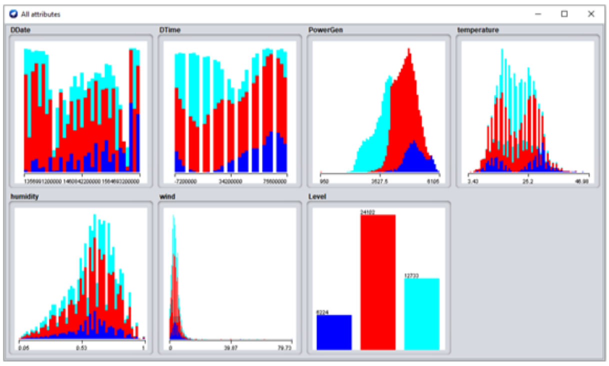

Figure 3. Visual depiction of the data

All twelve classifiers mentioned earlier in the evaluation technique are implemented using the real dataset "Load_Powe.arff". In the "Classify" tab, a classification algorithm is selected for implementation and the evaluation metrics are computed as given Figures 4-6.

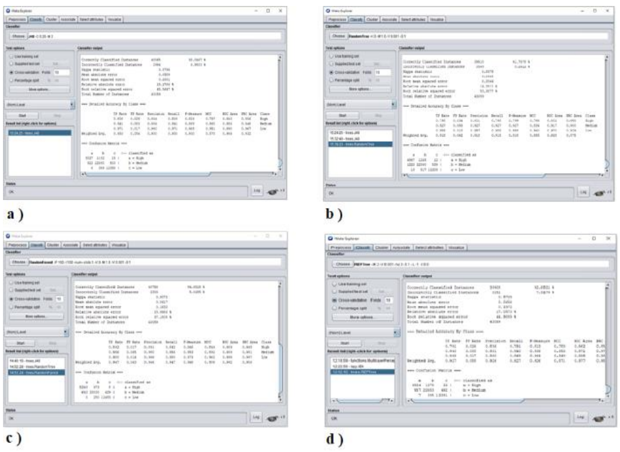

Figure 4. The evaluation metrics for: a) J48 Decision Tree algorithm, b) Random Tree algorithm, c) Random Forest algorithm and d) Decision Tree algorithm

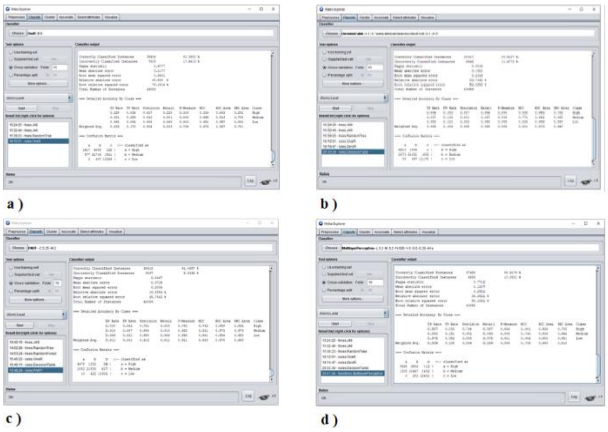

Figure 5. The evaluation metrics for: a) One Rule algorithm, b) Decision Table of Classification algorithm, c) Part of Classification algorithm and d) Multilayer Perceptron algorithm

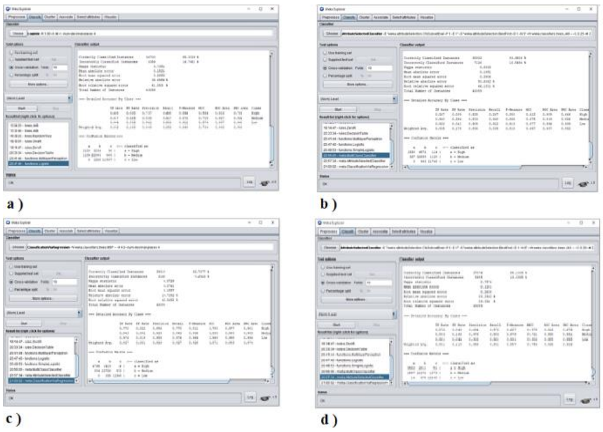

Figure 6. The evaluation metrics for: a) Logistic algorithm, b) Multi Class Classifier algorithm, c) Classification Via Regression algorithm and d) Attribute Selected Classifier algorithm

Table 2 shows the comparison of time taken to build the model and Performance of classifiers based on identification of correct instances with different algorithms.

Figure 7 shows the Random Tree, Decision Tree, and One Rule algorithms had the shortest training times, with build times of 0.39 seconds, 0.45 seconds, and 0.08 seconds, respectively. The Attribute Selected Classifier and Logistic algorithms also had relatively fast training times, with build times of 0.88 seconds and 0.89 seconds, respectively, Multi Class Classifier algorithm, J48, Decision Table algorithms had training times of 0.91 seconds, 1.25 seconds, and 3.99 seconds respectively, while the Classification Via Regression, Multilayer Perceptron, Part of Classification and Random Forest algorithms had comparatively longer training times, with build times of 4.28 seconds, 4.6 seconds, 5.71 seconds, and 9.56 seconds respectively.

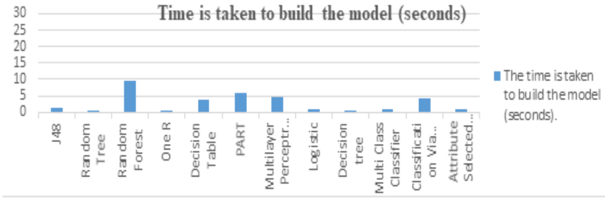

Figure 7. Comparison of time taken to build the model with different algorithms

In Figure 8, it looks like the Random Forest algorithm had the highest accuracy, with a correct identification rate of 94.65%. Followed by the J48 algorithm with a correct identification rate of 93.04%. The Random Tree algorithm and the Decision tree, Classification Via Regression, and Part of Classification algorithms also had relatively high accuracy, with correct identification rates of 91.75%, 92.68%, 92.70%, and 91.06%, respectively. Decision Table, Multilayer Perception, Attribute Selected Classifier and Logistic algorithms had good accuracy, with correct identification rates of 88.52%, 86.91%, 86.10% and 85.23%, respectively. Finally, Multi Class Classifier and One Rule algorithms had obtained correct identification rates of 83.45% and 82.33%, respectively. Figure 9 simply shows the percentage of correct classified and incorrect classified for each algorithm.

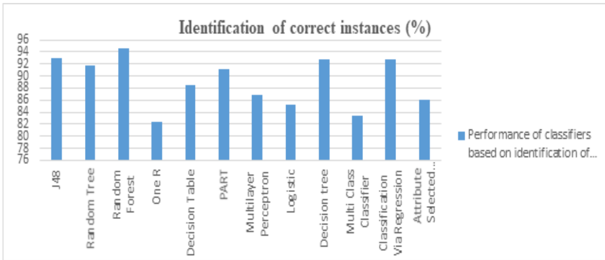

Figure 8. Comparison of Performance of classifiers based on identification of correct instances with different algorithm

Table 2. The time is taken (in seconds) to build the model and Performance of classifiers based on identification of correct instances

Classifier	The time taken to build the model (seconds).	Identification of correct instances (%)
J48	1.25	93.04
Random Tree	0.39	91.75
Random Forest	9.56	94.65
Decision tree	0.45	92.68
One Rule	0.08	82.33
Decision Table	3.99	88.52
Part of Classification	5.71	91.06
Multilayer Perceptron	4.6	86.91
Logistic	0.89	85.23
Multi Class Classifier	0.91	83.45
Classification Via Regression	4.28	92.70
Attribute Selected Classifier	0.88	86.10

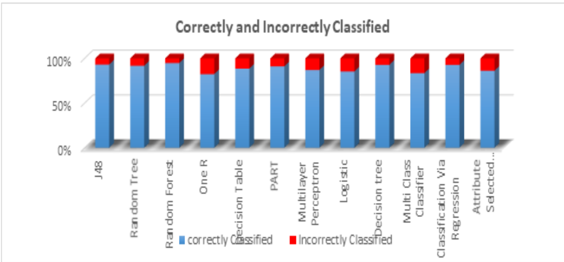

Figure 9. Correctly and incorrectly classifications

Based on the results for the classification algorithms on the energy data in this study, the Random Forest algorithm appears to be the best overall performer in terms of both accuracy and training time. It had the highest accuracy, with a correct identification rate of 94.651%, and a relatively fast training time of 9.56 seconds. Also the J48 and Classification Via Regression algorithms also had relatively high accuracy, with correct identification rates of 93.046% and 92.707%, respectively. These algorithms may be good options if accuracy is the primary concern. As for algorithms with fast training times, the One Rule, Random Tree, Decision Tree and Attribute Selected Classifier algorithms had built times of 0.08 seconds, 0.39 seconds, 0.45 seconds, and 0.88 seconds, respectively. These algorithms may be good options if speed is a key consideration.

3. CONCLUSION

In this study, twelve different classification methods are compared with each other for the analysis and estimation of energy consumption in Libya. The dataset used in this study is the historical data recorded from the General Electricity Company of Libya (GECOL) regarding the total electrical energy produced per hour in a certain period, temperature, humidity and wind speed recorded simultaneously. Especially, an artificial neural network with a multi-layer perceptron architecture is used by taking advantage of the predictive variable energy consumption. The input parameters of the neural network include power production, temperature, humidity and wind speed. Pre-processing of the SCADA (Supervisory Control and Data Acquisition) database was carried out to prepare the data for effective use in the neural network algorithm. A comprehensive set of experiments has been applied to build the prediction model, with evaluations based on accuracy and loss reduction. Artificial neural network is emerging as a valuable tool for energy prediction, with tests highlighting optimal performance at 1-fold using 6-fold cross-validation. In particular, an R2 value of 0.98 for learning data and 0.95 for testing data in 5-fold cross-validation demonstrates the effectiveness of the model.

In addition, the study uses the Apriori algorithm and Weka tool to analyze energy consumption models in detail. Twelve classification algorithms, namely J48, Random Tree, Random Forest, Decision Tree, One Rule, Decision Table of Classification, Part of Classification, Multilayer Perceptron, Logistic, Multi Class Classifier, Classification Via Regression and Attribute Selected Classifier, are implemented using the real dataset "Load_Powe.arff". Nine evaluation metrics, namely TP Rate, FP Rate, Precision, Recall, F-Measure, MCC, ROC Area, PRC Area and Class, are computed in order to compare the performances of twelve classification algorithms. In the field of energy data classification analysis, the Random Forest algorithm stands out as the best performing algorithm, achieving an impressive accuracy rate of 94.651% with a fast learning time of 9.56 seconds. Additionally, the J48 and Classification Via Regression algorithms achieve strong accuracy, achieving accuracy rates of 93.046% and 92.707%, respectively, suitable for applications that prioritize precision. For those emphasizing computational speed, alternatives such as One Rule, Random Tree, Decision Tree and Attribute Selected Classifier offer rapid training times ranging from 0.08 to 0.88 seconds. These results collectively contribute to a comprehensive understanding of energy management practices and offer valuable insights for future research and applications in the field.

ACKNOWLEDGEMENT

The authors would like to thank General Electricity Company of Libya (GECOL) for making the SCADA database available.

DECLARATION OF ETHICAL STANDARDS

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

AUTHORS' CONTRIBUTIONS

Ashraf Mohammed ABUSIDE: Applied the models, computed the metrics and analyzed the results.

Secil KARATAY: Applied the models, computed the metrics and organized the paper.

Rezvan REZAEIZADEH: Analyzed the results and organized the paper.

Aybaba HANCERLIOGULLARI: Analyzed the results and supervised and organized the paper.

CONFLICT OF INTEREST

There is no conflict of interest in this study

REFERENCES

[1] Bahssas D.M., AlBar A.M. and Hoque M.R., "Enterprise Resource Planning (ERP) Systems: Design, Trends and Deployment", *The International Technology **Management Review***, 5(2), 72 - 81, (2015).

[2] Alsuessi W., "General electricity company of Libya (GECOL)", ***European International Journal of Science and Technology***, 4(1): 61-69, (2015).

[3] Witten I.H., Frank E. and Hall M.A., "***Data Mining Practical Machine Learning Tools and Techniques***", Morgan Kaufmann, San Francisco, (2011).

[4] Abusida A.M. and Gultepe Y., "An Association Prediction Model: GECOL as a Case Study", ***International Journal of Information Technology and Computer Science***, 11(10): 34-39, (2019).

[5] Bouckaert R.R., Frank E., Hall M., Kirkby R., Reutemann P., Seewald A. and Scuse D., "***WEKA Manual for Version 3-6-10***", Hamilton, University of Waikato, /2013).

[6] Khamaira M.Y., Krzma A. and Alnass A.M., "Long Term Peak Load Forecasting for the Libyan Network", ***First Conference for Engineering Sciences and Technology (CEST-2018)***, Libya, (2018).

[7] Slimani T. and Lazzez A., "Efficient Analysis of Pattern and Association Rule Mining Approaches", ***International Journal of Information Technology and Computer Science***, 6(3): 70-81, (2014).

[8] Abusida A.M. and Hancerliogullari A., "A New Approach to Load Shedding Prediction in GECOL Using Deep Learning Neural Network", ***IJCSNS International Journal of Computer Science and Network Security***, 22(3): 220-228, (2022).

[9] Bank W., "***Supporting electricity sector reform in Libya: task C - institutional development and performance improvement of GECOL: report 4.2 - improving GECOL technical performance***", World Bank Group, Washington, (2017).

[10] Bhojani S. and Bhatt N., "Data Mining Techniques and Trends – A Review", *Global* ***Journal for Research Analysis***, 5(5): 252-254, (2016).

[11] Bharati M. and Ramageri B.M., "Data mining techniques and applications", ***Indian Journal of Computer Science and Engineering***, 1(4): 301-305 , (2010).

[12] Jain A.K., Murty M.N. and Flynn P.J., "Data Clustering: A Review", ***ACM Computing Surveys***, 31(3): 264-323, (1999).

[13] Zaki M.J. and Meira W., "***Data Mining and Machine Learning: Fundamental Concepts and Algorithms***", Cambridge University Press, New York, (2020).

[14] Bishop C.M., "***Pattern Recognition and Machine Learning (Information Science and Statistics)***", Springer, India, (2006).

[15] Hastie T., Tibshirani R. and Friedman J.H., "***The Elements of Statistical Learning Data Mining: Inference, and Prediction***", Springer,California, (2016).

[16] Duda R.O. and Hart P.E., "***Pattern Classification and Scene Analysis***", Wiley, California, (1973).