DEVELOPMENT OF AN INTELLIGENT SYSTEM FOR MODELING

Ozoh Patrick[1] , Ikechukwu Nwade[2], and Adepeju Adigun[1]

[1]Dept. of ICT, Osun State University

[2]Departemet of Mechanical Engineering, University of North Dakota

Abstract: To have insight into several model challenges, it is important to utilize artificial networks in models. Various research papers have been published about these artificial intelligence techniques. None of these publications addresses problems of the computation of estimates and forecasts for solving real-world data and models from estimated data.

Aims/Objectives: The objectives of this research are (1)To develop an artificial neural method for solving problems (2)The development of techniques to solve complex problems (3)The computation of estimates and forecasts for real-world data (4)The development of models from estimated data. The techniques investigated in this research are important and necessary for solving vague complex, and bogus problems in artificial intelligence. Consequently, a thorough study comprising techniques applied in this study is used for comparisons and utilized to identify a reliable method for modeling and forecasting problems. This research investigates different procedures utilized in artificial intelligence for modeling an efficient decision-making process.

Methodology/approach: Past studies of methods were utilized for this study. The methods applied in this paper include collecting data from REDcap, an online data collection tool. This was determined on the training dataset (70%) and evaluated on testing data (30%). The model is developed using the neural network, binary analysis, supervised learning classifier, and result determination.

Results/finding: The evaluation of results is done by comparing their performance using accuracy metrics. The model implementation was done using MATLAB programming language. The data was processed with an algorithm classifier.

Implication/impact: This work is advantageous in achieving efficiency in models. The artificial intelligence model is developed to improve the solution to issues in developing models.

Keywords: artificial intelligence, decision-making, forecasting, models

تطوير نظام الخبراء لعمليات المحاكاة

الملخص: للحصول على نظرة ثاقبة في العديد من تحديات النماذج ، من المهم استخدام الشبكات الاصطناعية في النماذج. تم نشر أوراق بحثية مختلفة حول تقنيات الذكاء الاصطناعي هذه. لا تتناول أي من هذه المنشورات مشاكل حساب التقديرات والتنبؤات لحل بيانات ونماذج العالم الحقيقي من البيانات المقدرة.

الأهداف / الأهداف: أهداف هذا البحث هي : (1) تطوير أسلوب عصبي اصطناعي لحل المشكلات (2) تطوير تقنيات لحل المشكلات المعقدة (3) حساب التقديرات والتنبؤات لبيانات العالم الحقيقي (4) تطوير النماذج من البيانات المقدرة. تعتبر التقنيات التي تمت دراستها في هذا البحث مهمة وضرورية لحل المشكلات المعقدة والغامضة المزيفة في الذكاء الاصطناعي. وبالتالي، يتم استخدام دراسة شاملة تشتمل على التقنيات المطبقة في هذه الدراسة للمقارنات واستخدامها لتحديد طريقة موثوقة للنمذجة والتنبؤ بالمشاكل. يبحث هذا البحث في الإجراءات المختلفة المستخدمة في الذكاء الاصطناعي لنمذجة عملية صنع القرار بكفاءة.

المنهجية / النهج: تم استخدام الدراسات السابقة للأساليب لهذه الدراسة. تتضمن الأساليب المطبقة في هذه الورقة جمع البيانات من REDcap ، وهي أداة لجمع البيانات عبر الإنترنت. تم تحديد ذلك في مجموعة بيانات التدريب (70٪) وتم تقييمه بناءً على بيانات الاختبار (30٪). تم تطوير النموذج باستخدام الشبكة العصبية والتحليل الثنائي ومصنف التعلم الخاضع للإشراف وتحديد النتيجة.

النتائج / النتائج: يتم تقييم النتائج بمقارنة أدائها باستخدام مقاييس الدقة. تم تنفيذ النموذج باستخدام لغة برمجة .MATLAB تمت معالجة البيانات باستخدام مصنف الخوارزمية.

التضمين / الأثر: هذا العمل مفيد في تحقيق الكفاءة في النماذج. تم تطوير نموذج الذكاء الاصطناعي لتحسين حل المشكلات في تطوير النماذج.

Corresponding author: patrick.ozoh@uniosun.edu.ng

1. Introduction

This research utilizes data from REDcap, an online data collection tool. The model is developed using the neural network, binary analysis, supervised learning classifier, and result determination techniques. Artificial intelligence involves discovering functions that can be performed without being directly programmed to do so. This consists of learning from available data to perform certain tasks. For other advanced functions, this can be challenging to produce the required algorithms. It can be more effective if the machine can develop its algorithm, as opposed to having answers.

Artificial intelligence consists of a wide variety of techniques utilized for the discovery of patterns and relationships in sets of information. The fundamental objective of any artificial intelligence algorithm is to recognize meaningful relationships in generating a generalization of these relationships that interpret new, unseen data. This is one applied aspect. [1] investigates sustainable innovation in businesses using various theoretical theorems. [2] improves the efficiency of models. It also improves the speed with which data are released. It is important to perform comparisons of models applied in machine learning [3]. These models can describe the relationships among the series. As a result, reliable decision-making is made.

The major focus of this study is to develop artificial intelligence models to enhance efficiency in models using artificial neural networks, binary analysis, supervised learning classifier, and the result determination utilizing MATLAB programming. This study has significant contributions. The summary of contributions from the study carried out includes the following:

1) The study gives a theoretical framework for artificial neural network models.

2) This research gives a systematic guide to producing a system consisting of an artificial intelligence system. As a result, the set of collected data is modeled with their normalization obtained during the estimation step is enhanced.

3) The procedure of controlling a system normalization assists in obtaining reliable estimates.

The background study of this research is contained in Section 1. The section is a suitable starting point for the study. Section 2 consists of a review of the literature of this study. This study will complement existing work. Section 3 discusses the theoretical framework of this study. It entails the prediction of data, including the algorithms for the techniques. Subsequently, this is applied to data samples drawn from REDcap, an online data collection tool. This section compares various techniques for estimating models. Section 4 estimates values for the various models. It selects the best technique by comparing their respective statistical estimates. Finally, Section 5 summarizes the research that has been accomplished and the findings of this study. This entails the evaluation of results done by comparing their performance using accuracy metrics. It also describes recommended future research direction of this study.

2. Literature Review

The necessity to utilize artificial intelligence techniques for modeling challenges was established in the previous section. The procedure and techniques used in this study are highlighted and discussed in this section. The theoretical guidelines of the techniques were established while the objectives, scope, and justification for this study were presented. One of the challenges of classifications in applying artificial intelligence is the use of suitable techniques to develop models dependent on available data. The model that would perform optimally among the data features is still open to research.

A variety of artificial learning techniques, despite their respective strengths and weaknesses, have been applied to enhance sustainability in organizations. [4] state their preparedness to apply artificial learning techniques in solving sustainability challenges in organizations that would promote their efficiency. The major objective for the application of artificial intelligence techniques is for encouraging the procurement of sustainable development goals. [5] discuss using data set from an online platform using kaggle.com. This is reviewed in this paper [6]. According to the paper, data can be divided into several groups namely; network system data, customer and supplier data, data collection, and consumer data. This is used to develop an incisive model of the distribution of class labels for predictor features. The learner is provided with two sets of data, a training set, and a test set. The supervised learning is displayed in Figure. 1.

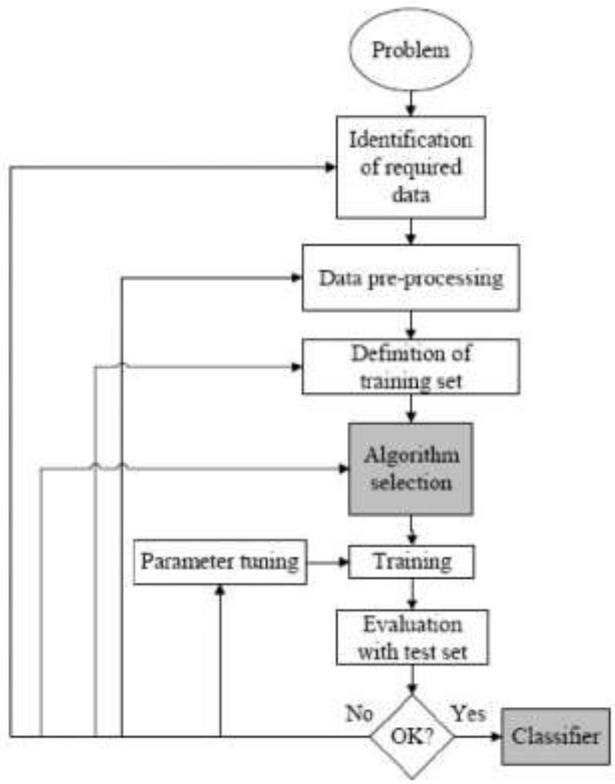

Figure 1. The processes

The research in [7] proposes an efficient framework for a time series analysis. A high-performance method is deployed to the application. The evaluation of the technique indicates that the technique is more accurate than previous techniques. The work in [8] develops efficient data-driven models for accurate forecasts. The aim of [9] is to propose data summarizing the structure of COVID. It utilizes Box Jenkins together with the linear exponential method. The results from the study show the combined technique to be more reliable.

Developing models for the environment is essential [10]. The study introduces an algorithm-based hybrid model. In the paper, the hybrid model performed better in forecasting. [11] proposes a method for developing observed data. In illustrating the evaluation process, some numerical problems were presented. [12] proposed surveying recent ANN applications. These techniques included models applied in machine learning. [13] reported the existence of similarities in models. These models apply to the same objects found in photographs, drawings, and images. [14] proposes the application of new models in residential load forecasting in neural network models. Content evaluation Is inferred using the proposed model compared to load forecasts in buildings.

[15] develops a model for reducing the spread of the COVID disease. A deterministic nonlinear incidence function was used to model the disease. Results from the study indicate the model will be maintained, otherwise, instability will occur. [16] uses chatbots. The study used electronic surveys that were distributed to SMEs operating in London to collect data between April and June 2020. This research applies a layout system to a sustainable business growth model [17]. According to [18], the evaluation and simulation of data were carried out by artificial intelligence, and the system was developed by presenting a method for identifying solutions to digitalization issues in organizations. [19] presents an enhanced set of data for collecting, processing, and controlling for efficiently handling the data produced for manufacturing procedures, which assists in the distribution processes and assessing the market in real-time. An application for an artificial intelligence platform is proposed by [20[. According to the paper, an attempt is made to present the need for quicker activities. The important characteristics for the efficient growth of industries are policies involved in their formulation.

[21] proposes that it is important for variables investigations to be done. These include variables that are important in size data that will affect growth in organizations. [22] investigates an organization’s intelligent financial construction

that will affect sustainability in the organizational growth process. This is to provide for organizations requiring research. Moreover, [23] consists of several journals on different sets of challenges facing sustainable development. Furthermore, this paper provides a new understanding of the importance of artificial intelligence applications and models, deterministic nonlinear incidence function was used to model the disease. Results from the study indicate the strength of the model will be maintained, otherwise, instability will occur.

3. Theoretical Framework

The techniques used in this study are presented in this section. This section presents techniques utilized in this study. This section includes data collection, in which data is collected from an online data collection source. The system development is for the application of machine learning techniques undertaken by implementing various machine learning algorithms using MATLAB programming language to input data sets via the artificial neural network, binary analysis, supervised learning classifier, and result determination techniques.

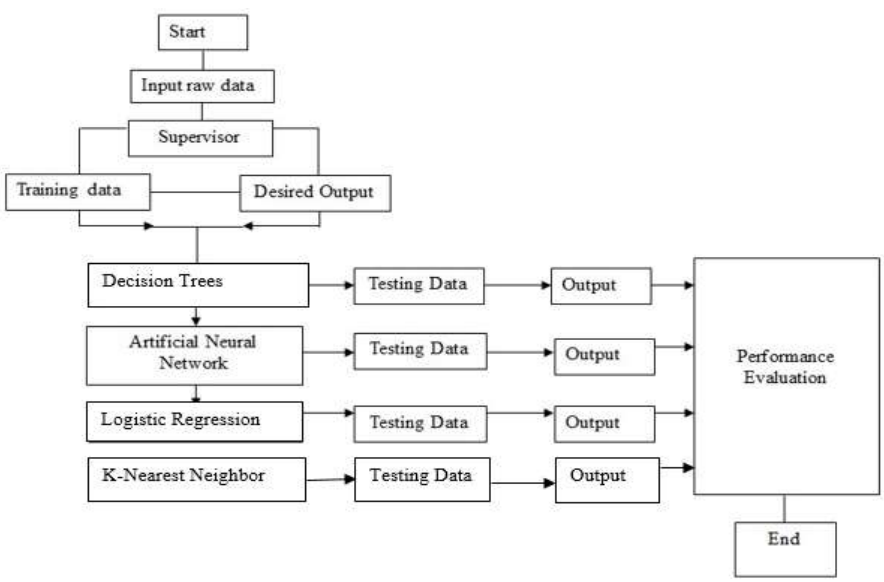

Figure 2. Comparative analysis model.

The ability of these techniques is done using an evaluative mechanism for performance testing. The methodology for the classification is displayed in Figure 2.

Several of the data sets are taken as input for the feature extractor and classification algorithm. The datasets are entered into a sequence of pre-processing blocks. Achieving this objective involves the steps that are depicted in Figure 3. The MATLAB programming language is used to evaluate the models developed in this study. The MATLAB code used for evaluation is given in Appendix 5.

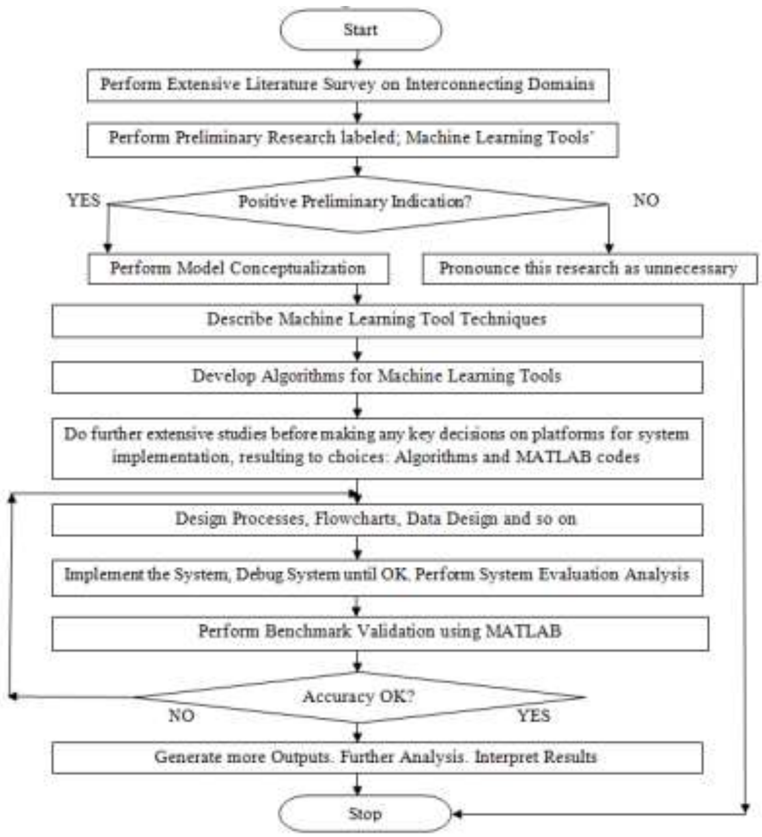

Figure 3. Flowchart of methodology.

The techniques used in this research are presented.

3.1 Data Set

All The set of data is collected from the REDcap online data collection application. REDcap is an online data collection application [24]. A sample of 22 companies was taken with their respective data containing company category, employees, turnover, and balance sheet. It is given in Figure 4.

Company Category	Employees	Turnover	Balance sheet total
Micro	<10	<€2 million	<€2 million
Small	<50	<€10 million	<€10 million
Medium sized	<250	<€50 million	<€43 million

Figure 4. Sample data.

3.2 Artificial Neural Network

ANNs were created for defining the architecture of humans to perform functions that standards can not do. The neural network is classified into a training set, validation set, and evaluation. They provide important techniques for creating systems used in building models [25]. This is shown in Figure 5. Appendix 1 shows the code for Artificial Neural Networks.

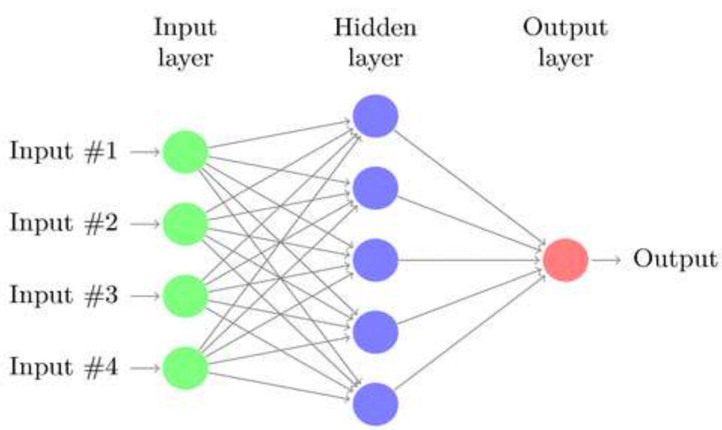

Figure 5. Model of an artificial neural network.

3.3 Binary Analysis

Binary analysis is a statistical system ensuring that its standard form uses a logistic function to model a binary-dependent function, although several complex extensions exist. The binary analysis model is given in Figure 6. The characteristics of the binary analysis are given in Figure 7.

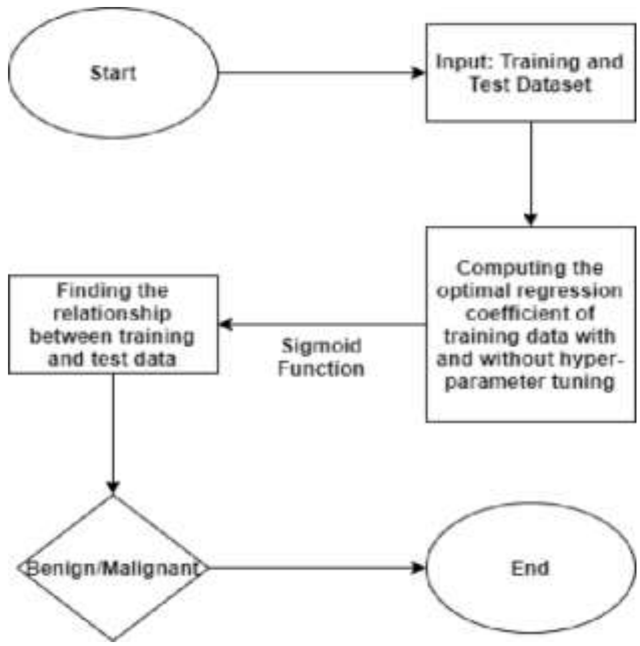

Figure 7. Estimates of binary analysis.

The mathematical notion of binary analysis consists of the variable and the independent variables. The binary analysis model can be written as:

$$(Y) = \text{In}\ (1-\pi) = \beta_0 + \beta_1 X \quad (1)$$

β_0 and β_1 are defined as intercept and regression coefficients respectively.

The antilog on both results in

$$\pi = P(Y|X = x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \quad (2)$$

Binary analysis can entail more than one predictor

$$(Y) = In\ 1{-}\pi{=}\ \beta_0 + \beta_1 + \cdots + \beta_p X_p \qquad (3)$$

Equation (3) is the standard form of a binary analysis model for p predictors.

The regression parameter β is estimated by the maximum likelihood (ML) method. The value of regression coefficients β1...βp corresponds between X and the logit of Y. The odds ratio is estimated using Equation (4)

$$Odd\ (OR) = e^{\beta} \qquad (4)$$

Appendix 2 depicts the MATLAB code used to implement the binary analysis.

3.4 Supervised Learning Classifier

The supervised learning classifier comprises the k-closest training and the output. The results consist of the property value. supervised learning classifier with an algorithm since the data assigned to it is labeled. It is a non-parametric technique as the classification of a test data point depends on the nearest training data points comparatively to considering the variables of the dataset. It is used in resolving classification and regression tasks. Supervised learning utilizes training data to learn the mapping variable that converts input variables (X) into the output variable (Y). It solves for *f* given equation (5).

$$Y = f\ (X) \qquad (5)$$

It allows the accurate generation of outputs given new inputs.

There are different types. They are classification, regression, and the ensemble.

Classification is predicting results in the description of categorical data.

Regression is used to predict the data that can be described as real values.

An ensemble is a form of supervised learning. It is integrating the predictions of several machine learning models that are separately weak to get an accurate prediction on a new sample. Appendix 3 shows the MATLAB code for the supervised learning classifier.

3.5 Result Determination Technique

A result determination technique consists of a flowchart-like structure, with each node constituted by an attribute and each branch composed of the results of the test, and an individual node consisting of a class label. A result determination consists of chance event outputs and resource costs. The directions from the root to the leaf composes of classification rules. The result determination technique process is given as follows:

1. Decision nodes – composed of squares
2. Chance nodes – comprises of circles
3. End nodes – represented by triangles

The result determination s algorithm comprises two parts: nodes and rules. The standard suggestion of the algorithm is to produce a flowchart diagram that comprises a root node on top. The important factors why the result determination technique is used in data mining and classification. Figure 8 represents a result determination algorithm.

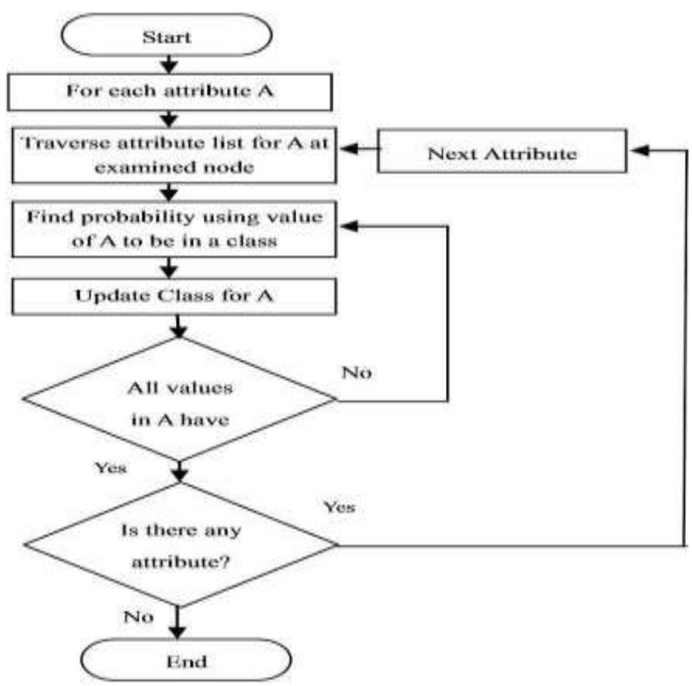

Figure. 8 Result determination algorithm.

Appendix 4 depicts the MATLAB code used to implement the result determination model. The prediction for this system is implemented using MATLAB and evaluated using the mean squared error.

3.6 Evaluation of Techniques

There are several techniques used for the evaluation of results. The mean squared error (MSE) technique is an accuracy metric used to evaluate the performance of the results in this study. This accuracy metric is the most commonly used in past literature. The comparative analysis of the techniques involves identifying and selecting a reliable and accurate technique by comparing the *mean square error* (MSE) for the respective models, as discussed by [26]. Reliable and accurate techniques must be employed in the modeling of the models, otherwise poor estimates may result. The performances of the neural network, binary analysis, supervised learning classifier, and result determination is assessed by comparing their respective *MSE* values. The technique with the smallest error value is selected as the best-performing technique as discussed by [27].

This approach of assessing the performance of a proposed technique as compared with benchmarks was also used by [28]. The error estimates for the four techniques used to model data in this study are provided to compare the performance of the models and their reliability by comparing actual data with predicted models. *The MSE* values were described by [29] and are outlined as follows:

The errors are computed from a time series, based on an average of weighted past observations. At period *t*, past values of a variable of interest Xt can be observed. The model is applied to the historical observations, and the values Ft+1 are obtained. Once the values are obtained, they can be compared with known values and the error *e* can be calculated. To identify an accurate model, the following steps are followed:

- Choose a method based on previous knowledge about the observed pattern of the time series.
- Use the method to develop fitted values of the data.
- Calculate the error.
- Decide on the appropriateness of the model based on the measure of the error

For the purpose of computing errors, a historical data set called a time series is considered. The time series is represented as follows:

$$F_{t+1} = \frac{X_t + X_{t-1} + \cdots + X_{t-n+1}}{n} \quad (6)$$

$$= \frac{1}{n}\left(\sum_{i=t-n+1}^{t} X_i\right) \quad (7)$$

where t is the most recent observation and $t + 1$ is the next period.

An examination of the error permits the evaluation of whether the chosen model accurately mirrors the pattern exhibited in the sample observations. An evaluation of the reliability of a model requires the specification of criteria.

This study utilizes the *mean square error* (MSE) technique to evaluate the accuracy of the models. The *MSE* is given as

$$MSE = \frac{\sum(e_t)^2}{n} = \frac{\sum(X_t - F_t)^2}{n} \quad (8)$$

This measure defines error as the sum of et divided by the sample size, that is, the number of periods. The *MSE* is a measure of individual errors.

4. Results

This study includes the implementation framework and results of the programming language used from the preprocessing phase to the training and also validation phase of the prediction models. Screenshots of results are presented to support the proposed framework. MATLAB software is used to analyze the techniques used in this study. Appendix 6 shows the MATLAB code used to read the data set file. The data is split into train and test groups. For classification, the preprocessed data is fed to the binary analysis model. The data is split into two parts; the training set (70% of data) and the test set (30% of data). The test set is used to evaluate the performance of the model. Appendix 7 shows the MATLAB programming code used to implement this.

Appendix 7 depicts the MATLAB code used to implement the training of the supervised learning classifier. Appendix 8 depicts the steps for the supervised learning classifier. Appendix 9 depicts the steps for the decision tree model. The test data is used to test the intelligence of the model. The test set is fed to the model for the model to make predictions. The predicted outcomes from the model will be used to match the actual outcomes of the test set. After the training phase, the classifier is tested and its prediction accuracy is measured. To effectively evaluate the performance of these prediction models, the evaluation criteria used is the mean squared error. The MSE either assesses the quality of a predictor. The MSE of the predictor is computed as

$$\mathbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

The output for the MSE of selected technique is given in Figure 9. In describing Figure 9, the data for the REDcap online data is analyzed. The data is shown in the graph analysis by MATLAB software, as shown in the figure. The analysis selects the best method for this study. The neural network is selected. Figure 9 shows that the MSE of the neural network reduces over time. The graph also shows the time, performance, and validity check of the analysis. The execution time is very efficient. It is given as 0.0001. As given in the figure, the validity check is good when estimates are compared with actual data.

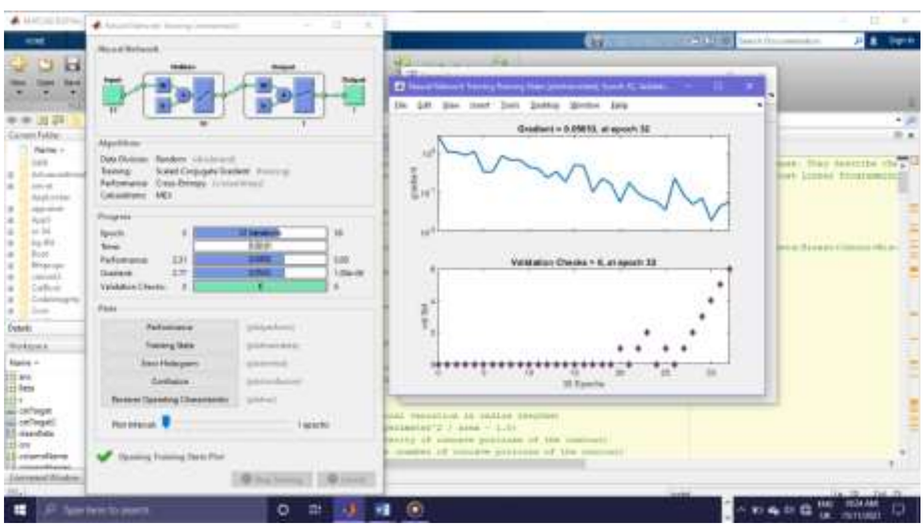

Figure 9 MSE model.

The output for the validity performance is given in Figure 10. The figure shows the validity performance of the neural network. The performance shows a reducing curve, which indicates efficiency. The graph also shows a reduction in test and trained data. The sharp decrease in the graph signifies a reduction in the MSE and validity values. Overall, there was a reducing trend in data. This shows that data has been following a reducing trend over time.

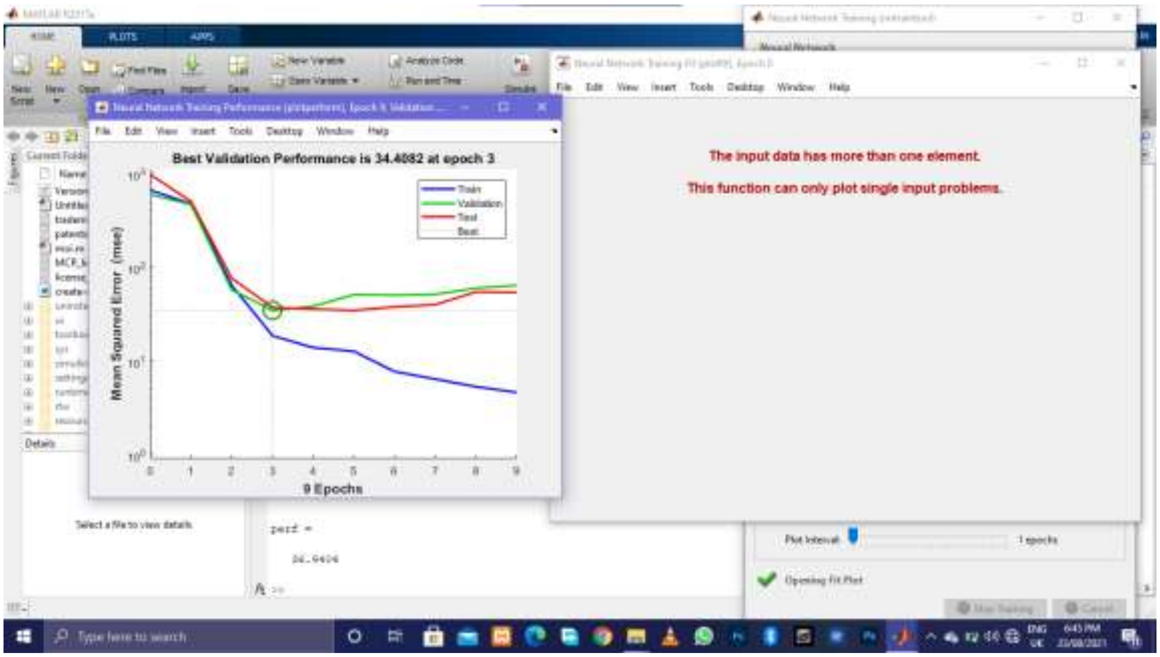

Figure 10. Validity Performance.

When the evaluation techniques are implemented on the REDcap online data collection application, the MSE of the proposed technique, the neural network has the lowest mean squared error of 0.02006, compared with the supervised learning classifier (0.02529), result determination (0.06471) and the binary analysis (0.02941).

This is shown in Table 1. As a result, the neural network has the best prediction accuracy and the lowest mean squared error at 0.02006.

Table 1. Results of MSE for Techniques

Algorithm	MSE
Artificial Neural Network	0.02006
Supervised Learning Classifier	0.02529

Result Determination	0.06471
Binary Analysis	0.02941

Figure 11 and Figure 12 are not required as the study requires the MSE and the validity of the technique.

This study also compares results obtained with previous similar research. This study compares results with [30]. The results for the error values of both studies are given in Table 2.

ANN error value for this study	ANN error value for [30]
0.02006	0.355216

The error value of this study is significantly lower than the previous study. This indicates this study is more efficient than the previous study.

5. Conclusion

A variety of artificial learning techniques, despite their various strengths and Weaknesses were applied in this study. This study compares techniques to identify their reliability. compares various artificial learning techniques to identify a reliable method. The testing of the techniques is focused on 22 testing data datasets. Results show that ANN is the most accurate for estimating data. The proposed method has improved previous methods of modeling data. The techniques are evaluated by investigating the accuracy of the predicted values for artificial neural network (ANN), supervised learning classifier, binary analysis, and result determination techniques by computing and comparing their respective Mean Squared Error (MSE). To identify a robust technique, performance measures are used to test the accuracy of the developed models. This is to investigate the robustness of the techniques. The MATLAB code for evaluation is given in Appendix 5.

The recommended future research for this study involves the following process. Currently, applying artificial intelligence techniques to solving issues with modeling can be important research. These challenges include variables contained in models. Those issues can improve the performance of applying artificial intelligence techniques to the models and getting to know the contributions of each of these variables to the models will need further research.

6. References

[1] Govindan, Kannan. How artificial intelligence drives sustainable frugal innovation: A multitheoretical perspective. IEEE Transactions on Engineering Management (2022).

[2] Squazzoni, Flaminio, J. Gareth Polhill, Bruce Edmonds, Petra Ahrweiler, Patrycja Antosz, Geeske Scholz, Emile Chappin et al. Computational models that matter during a global pandemic outbreak: A call to action. (2020).

[3] Ozoh, Patrick, Moruf Olayiwola, and Adepeju Adigun. An In-Depth Study of Typical Machine Learning Methods via Computational Techniques. Analele Universităţii "Tibiscus", Timişoara 16, no. 2 (2018): 77-81.

[4] Palomares, Iván, Eugenio Martínez-Cámara, Rosana Montes, Pablo García-Moral, Manuel Chiachio, Juan Chiachio, Sergio Alonso et al. A panoramic view and swot analysis of artificial intelligence for achieving the sustainable development goals by 2030: Progress and prospects. Applied Intelligence 51, no. 9 (2021): 6497-6527.

[5] Mansur Huang, Nur Shahellin, Zaidah Ibrahim, and Norizan Mat Diah. "Machine learning techniques for early heart failure prediction." Malaysian Journal of Computing (MJoC) 6, no. 2 (2021): 872-884.

[6] Ahmad, Tanveer, Dongdong Zhang, Chao Huang, Hongcai Zhang, Ningyi Dai, Yonghua Song, and Huanxin Chen. Artificial intelligence in sustainable energy industry: Status Quo, challenges and opportunities. Journal of Cleaner Production 289 (2021): 125834.

[7] Fernandez, Ivan, Ricardo Quislant, Sonia Gonzalez-Navarro, Eladio Gutierrez, and Oscar Plata. TraTSA: A Transprecision Framework for Efficient Time Series Analysis. Journal of Computational Science 63 (2022): 101784.
[8] Alkesaiberi, Abdulelah, Fouzi Harrou, and Ying Sun. Efficient wind power prediction using machine learning methods: A comparative study. Energies 15, no. 7 (2022): 2327.
[9] Yonar, Harun, Aynur Yonar, Mustafa Agah Tekindal, and Melike Tekindal. Modeling and Forecasting for the number of cases of the COVID-19 pandemic with the Curve Estimation Models, the Box-Jenkins and Exponential Smoothing Methods. EJMO 4, no. 2 (2020): 160-165.
[10] Kilinc, Huseyin Cagan, and Adem Yurtsever. Short-Term Streamflow Forecasting Using Hybrid Deep Learning Model Based on Grey Wolf Algorithm for Hydrological Time Series. Sustainability 14, no. 6 (2022): 3352.
[11] Ye, Tingqing, and Baoding Liu. Uncertain hypothesis test with application to uncertain regression analysis. Fuzzy Optimization and Decision Making 21, no. 2 (2022): 157-174.
[12] Yang, Xu, Jinchao Guan, Ling Ding, Zhanping You, Vincent CS Lee, Mohd Rosli Mohd Hasan, and Xiaoyun Cheng. Research and applications of artificial neural network in pavement engineering: a state-of-the-art review. Journal of Traffic and Transportation Engineering (English Edition) 8, no. 6 (2021): 1000-1021.
[13] Goh, Gabriel, Nick Cammarata, Chelsea Voss, Shan Carter, Michael Petrov, Ludwig Schubert, Alec Radford, and Chris Olah. Multimodal neurons in artificial neural networks. Distill 6, no. 3 (2021): e30.
[14] Panda, Saroj Kumar, and Papia Ray. Analysis and evaluation of two short-term load forecasting techniques. International Journal of Emerging Electric Power Systems 23, no. 2 (2022): 183-196.
[15] Odetunde, Opeyemi, Mohammed Olanrewaju Ibrahim, Olanrewaju Thomas Olotu, Ohigweren Airenoni Uwaheren, and Tosin Akinyemi. Stability and bifurcation analysis of COVID-19 mathematical model incorporating case detection. Malaysian Journal of Computing (MJoC) 7, no. 1 (2022): 952-967.

[16] Drydakis, Nick. Artificial Intelligence and reduced SMEs' business risks. A dynamic capabilities analysis during the COVID-19 pandemic. Information Systems Frontiers (2022): 1-25.

[17] Wong, Lai-Wan, Garry Wei-Han Tan, Keng-Boon Ooi, Binshan Lin, and Yogesh K. Dwivedi. Artificial intelligence-driven risk management for enhancing supply chain agility: A deep-learning-based dual-stage PLS-SEM-ANN analysis. International Journal of Production Research (2022): 1-21.

[18] Barton, Martin, Roman Budjac, Pavol Tanuska, Gabriel Gaspar, and Peter Schreiber. Identification Overview of Industry 4.0 Essential Attributes and Resource-Limited Embedded Artificial-Intelligence-of-Things Devices for Small and Medium-Sized Enterprises. Applied Sciences 12, no. 11 (2022): 5672.

[19] Manimuthu, Arunmozhi, V. G. Venkatesh, Yangyan Shi, V. Raja Sreedharan, and SC Lenny Koh. Design and development of automobile assembly model using federated artificial intelligence with smart contract. International Journal of Production Research 60, no. 1 (2022): 111-135.

[20] Sharma, Paawan, Jigarkumar Shah, and Reema Patel. Artificial intelligence framework for MSME sectors with focus on design and manufacturing industries. Materials Today: Proceedings (2022).

[21] Helmer, Pernilla, Joseph Halim, Jie Zhou, Roopathy Mohan, Björn Wickman, Jonas Björk, and Johanna Rosen. Investigation of 2D Boridene from First Principles and Experiments. Advanced Functional Materials 32, no. 14 (2022): 2109060.

[22] Yang, Ning. Financial Big Data Management and Control and Artificial Intelligence Analysis Method Based on Data Mining Technology. Wireless Communications and Mobile Computing 2022 (2022).

[23] Musleh Al-Sartawi, Abdalmuttaleb MA, Khaled Hussainey, and Anjum Razzaque. The role of artificial intelligence in sustainable finance. Journal of Sustainable Finance & Investment (2022): 1-6.

[24] Willers, Craig, Tom Lynch, Vibhasha Chand, Mohammad Islam, Marissa Lassere, and Lyn March. A versatile, secure, and sustainable all-in-one biobank-registry data solution: The A3BC REDCap model. Biopreservation and biobanking 20, no. 3 (2022): 244-259.

[25] Rahman, Md Atiqur, and Yang Wang. Optimizing intersection-over-union in deep neural networks for image segmentation. In International symposium on visual computing, pp. 234-244. Springer, Cham, 2016.
[26] Hodson, Timothy O. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geoscientific Model Development 15, no. 14 (2022): 5481-5487.
[27] Adame, Berhan Oumer, and Ayodeji Olalekan Salau. Genetic algorithm based optimum finger selection for adaptive minimum mean square error rake receivers discrete sequence-CDMA ultra-wide band systems. Wireless Personal Communications 123, no. 2 (2022): 1537-1551.
[28] Ćalasan, Martin, Shady HE Abdel Aleem, and Ahmed F. Zobaa. On the root mean square error (RMSE) calculation for parameter estimation of photovoltaic models: A novel exact analytical solution based on Lambert W function. Energy conversion and management 210 (2020): 112716.
[29] Han, Yongming, Jingze Li, Xiaoyi Lou, Chenyu Fan, and Zhiqiang Geng. "Energy saving of buildings for reducing carbon dioxide emissions using novel dendrite net integrated adaptive mean square gradient." Applied Energy 309 (2022): 118409.
[30] Ozoh, Patrick, Morufu Oyedunsi Olayiwola, and Adepeju Abeke Adigun. Prediction of electricity consumtion based on complex computational method. Mathematical Sciences and Informatics Journal (MIJ) 3, no. 1 (2022): 56-65.

Appendix 1

MATLAB Code for Training the ANN model

```
%% TEST NEURAL NETWORK CLASSIFIER %%
dataTestMat=transpose(dataTest{:,:});
targetTestMat=transpose(targetTest{:,:});

testResult = net(dataTestMat);
testIndices = vec2ind(testResult);
subplot(2,2,2);
plotconfusion(targetTestMat,testResult)
[c,cm] = confusion(targetTestMat,testResult);
fprintf('Percentage Correct Classification   : %f%%\n', 100*(1-c));
fprintf('Percentage Incorrect Classification : %f%%\n', 100*c);
subplot(2,2,3);
plotroc(targetTestMat,testResult)
% perf = mse(net, targetTestMat,testResult ,'regularization', 0.01)
err = immse(targetTestMat,testResult);
fprintf('mse of Neural network is: %f\n' ,err)
nnErr=err;
```

Appendix 2
MATLAB Code for binary analysis

```
%% training set %%
mtrain=length(categorical(targetTrain{:,:}));
dataTrainMat2=[ones(mtrain,1) transpose(dataTrainMat)];
ztrain = dataTrainMat2*Beta;%z=beta0*1+beta1*x1+beta2*x2
htrain=1.0./(1.0+exp(-ztrain));
figure
subplot(2,1,1);
histogram(htrain,10)
xlabel('Probabilities')
ylabel('Frequencies')
title('Probability Distribution for Training Data')
```

Appendix 3
MATLAB Code for Training the KNN Model

```
%% KNN %%
knnModel=fitcknn(dataTrain{:,:},categorical(targetTrain{:,:}),'NumNeighbors',10,'Standardize',1);
disp(knnModel);
predictedValues=predict(knnModel,transpose(dataTestMat));
pred = grp2idx(predictedValues)-1
err = immse(targetTestMat,transpose(pred));
fprintf('mse of KNN is: %f\n' ,err)
knnErr=err;
```

Appendix 4
MATLAB Code for Training the DT Model

```
%% DECISION TREES %%
decisionTreesModel=fitctree(dataTrain{:,:},categorical(targetTrain{:,:}));
disp(decisionTreesModel);
view(decisionTreesModel,'Mode','graph')
predictedValues=predict(decisionTreesModel,transpose(dataTestMat));
pred = grp2idx(predictedValues)-1
err = immse(targetTestMat,transpose(pred));
fprintf('mse of decision Trees is: %f\n' ,err)
dtErr=err;
```

Appendix 5
MATLAB Code for Evaluation

```
function V=errperf(T,P,M)
%   rmse (root mean squared error)
%   mape (mean absolute percentage error)
%   e (errors)
%   se (squared errors)
%{
Abbreviations:
e: error(s)
M: METRIC
m: mean
P: PREDICTIONS
p: percentage
s: squared
T: TARGETS
V: VALUE(S)
%}
% Transform input
M=lower(M);
%% Compute metric
switch M
% Errors
case 'e'
  V=T-P;
% Squared errors
case 'se'
  Ve=errperf(T,P,'e');
  V=Ve.^2;
% Mean squared error
case 'mse'
  Vse=errperf(T,P,'se');
  V=mean(Vse);
case 'rmse'
  Vmse=errperf(T,P,'mse');
  V=sqrt(Vmse);
% Relative errors
case 're'
  assert(all(T),'All elements of T must be nonzero.')
  Ve=errperf(T,P,'e');
  V=Ve./T;
% Percentage errors
case 'pe'
  Vre=errperf(T,P,'re');
  V=Vre*100;
% Absolute percentage errors
case 'ape'
  Vpe=errperf(T,P,'pe');
          V=abs(Vpe);
        % Mean absolute percentage error
          case 'mape'
          Vape=errperf(T,P,'ape');
          V=mean(Vape);
```

Appendix 6
MATLAB Code for data set file

```
%% TABLE HEATMAP %%
inputCorr=corr(cleanData{:,:});
imagesc(inputCorr);
newColumns={};
columnNames = cleanData.Properties.VariableNames
for columnName = columnNames
  disp(['the member is ' columnName{1}])
  name= columnName{1};
  if contains(name,'_')
    newColumns(end+1)={strrep(name,'_','-')};
  else
    newColumns(end+1)={name};
  end
end
[X,Y]=meshgrid(1:31,1:31);
set(gca, 'XTick', 1:width(cleanData)); % center x-axis ticks on bins
set(gca, 'YTick', 1:width(cleanData)); % center y-axis ticks on bins
set(gca, 'XTickLabel', newColumns); % set x-axis labels
set(gca, 'YTickLabel', newColumns); % set y-axis labels
set(gca,'FontSize',10) % Creates an axes and sets its FontSize to 18
xtickangle(45)
title('Correlation Heat MAP', 'FontSize', 10); % set title
colormap('jet'); % Choose jet or any other color scheme
txt=sprintfc('%.1f',inputCorr)
text(X(:),Y(:),txt,'horizontalalignment','center','verticalalignment','middle')
colorbar()
```

Appendix 7
MATLAB Code for training of the supervised learning classifier

```
%% TEST-TRAIN DATA SPLIT %%

cv = cvpartition(size(cleanData,1),'HoldOut',0.3);
idx = cv.test;
% Separate to training and test data
dataTrain = cleanData(~idx,:);
dataTest  = cleanData(idx,:);
m = transpose(dataTrain.id);
rowsidx_train = any((targetDataWithId.id==m),2)
targetTrain=targetDataWithId(rowsidx_train,:)
k = transpose(dataTest.id);
rowsidx_test = any((targetDataWithId.id==k),2)
targetTest=targetDataWithId(rowsidx_test,:)
size(targetTest)
size(targetTrain)
catTarget2 = categorical(targetTrain{:,2});
countTarget2 = countcats(catTarget2);
subplot(2,1,2);
h=histogram(catTarget2)
ylabel('Number of Cases')
title('Histogram of Training Data Categories')
%h = histogram('Diagnosis',catTarget,'Counts',2);
y = h.Values;
text(1:2, y+5, string(y));
```

Appendix 8
MATLAB Code for the supervised learning classifier

```
%% TEST NEURAL NETWORK CLASSIFIER %%
dataTestMat=transpose(dataTest{:,:});
targetTestMat=transpose(targetTest{:,:});

testResult = net(dataTestMat);
testIndices = vec2ind(testResult);
subplot(2,2,2);
plotconfusion(targetTestMat,testResult)
[c,cm] = confusion(targetTestMat,testResult);
fprintf('Percentage Correct Classification   : %f%%\n', 100*(1-c));
fprintf('Percentage Incorrect Classification : %f%%\n', 100*c);
subplot(2,2,3);
plotroc(targetTestMat,testResult)
% perf = mse(net, targetTestMat,testResult ,'regularization', 0.01)
err = immse(targetTestMat,testResult);
fprintf('mse of Neural network is: %f\n' ,err)
nnErr=err;
```

Appendix 9
MATLAB Code for decision tree model

```
%% DECISION TREES %%
decisionTreesModel=fitctree(dataTrain{:,:},categorical(targetTrain{:,:}));
disp(decisionTreesModel);
view(decisionTreesModel,'Mode','graph')
predictedValues=predict(decisionTreesModel,transpose(dataTestMat));
pred = grp2idx(predictedValues)-1
err = immse(targetTestMat,transpose(pred));
fprintf('mse of decision Trees is: %f\n' ,err)
dtErr=err;
```