



PERFORMANCE EVALUATION OF NARX, RF AND LR MODELS FOR PREDICTION OF MEASLES DISEASE

KULUWA HAUWA AHMAD, JOHN KOLO ALHASSAN & ABDULLAHI IBRAHIM MOHAMMED

Computer Science Department, Federal University of Technology Minna

Abstract

This work is on Performance Evaluation of Nonlinear Autoregressive Recurrent Neural Networks with exogenous input (NARX), Random Forest (RF) and Linear

Keywords:
Data Mining techniques, Measles disease, MATLAB, and Predictive model. Regression (LR) for prediction of measles disease. Predicting measles disease is a difficult task due to

INTRODUCTION

Due to several seasonable time changes, people get affected by more and more contagious diseases; these diseases can be predicted in advance using prediction models (Sumathi et al., 2016). Diseases are the abnormal medical conditions of organisms that impair bodily functions, and are associated with recognizable symptoms and signs (Idowu et al., 2013). The causes of diseases may be related to external factors such as infectious disease or autoimmune disease in the case of internal dysfunctions (Clark, 2007). Ability to predict the future diseases enables clinicians and other

seasonable time studied by exploring
changes of the disease other machine
rate that vary between learning models aside
different locations. The NARX NN, RF, and LR.

NARX, RF, and LR models were used to predict Measles for the data collected from Federal medical center, Bida, Niger State, Nigeria, and their performance were compared. The results obtained for predicting measles showed that the NARX model proved to be most accurate because it had smaller RMSE of 6.7483 when compared with the RF of 14.4463 and LR of 23.6065. Therefore, this paper argues that using this model would enhance the effectiveness of routine immunization in Nigeria. The proposed model is recommended for usage by the researchers and clinicians. Some other diseases can be

Commented [-1]: This statement does not capture the problem, so rephrase

Commented [-2]: Add the specific values obtained for each model

Commented [ka3]:

Commented [-4]: Am not sure if you did this

researchers to take the most appropriate decisions. In traditional healthcare environments, diagnosis of a disease depends on doctor's decision for identifying it as the most likely cause depending on a person's symptoms. However, this leads to unwanted errors that resulting on more medical costs and affecting the quality of service provided to patients. There are various techniques used for the prediction of diseases. Among the various techniques, data mining plays an important role in the medical field.

Data Mining is defined as procedure of extracting information from huge sets of data (Vashi et al., 2017). There is a huge amount of data available in the Information Industry and this data is of no use until it is converted into useful information. Extraction of information is not the only process involved, it also includes other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation. Once all these processes are over, the information could be used in many applications such as Fraud Detection, Market Analysis, Production Control, Science Exploration, Medical Prediction etc. Data mining is significantly applied to medicine for the diagnosis of several diseases such as skin cancer, breast cancer, lung cancer, diabetes, liver disorder, heart disease, kidney failure, kidney stone, Hepatitis etc. (Kautish and Ahmed, 2016).

This research work focuses on development and assessment of Artificial Neural Network (ANN) model based on Non- Linear Autoregressive neural network with exogenous input (NARX), Linear Regression (LR), and Random Forest (RF) as well as the application of these models to the problem of measles prediction. Thus, the purpose is to present a triangulation of these models and as a result, reports a Mathematical Model with a savvy of these models for prediction using Federal Medical Center Bida (FMCB) as a case study.

LITERATURE REVIEW

Data mining has been evolving gradually in almost every area of life. A major area where Data Mining has made great strides is Education as explored in (Baker & Yacef, 2009). Further developments in the evolution

of Data Mining in Education is seen in (Siemens, 2013). In order to understand more about teaching, learning, intelligent contents and adaptation, several organizations (mostly members of Society for Learning Analytics Research and the International Educational Data Mining Society) came together to carry out numerous research on Learning Analytics and to find out the roles of Data Mining in Education. According to the paper, researches have also been done on Learning Analytics and how its applications in organizations, institutions and companies; in the paper it was stated that through journals, conferences, research labs and summer institutes the field has significantly grown in importance.

Prediction of Measles using Machine Learning Models

Measles is a highly contagious and dangerous disease caused by a virus. It only affects humans and rapidly spreads among individuals who have not been vaccinated. Measles is one of the most readily transmitted communicable diseases and arguably the most widely known and most deadly of all childhood rash/fever illnesses. Despite the availability of a safe and effective vaccine, Measles remains one of the leading causes of death among young children worldwide.

It is a common knowledge that climate plays a role in the transmission of many infectious diseases, some of which are among the most important causes of mortality and morbidity in developing countries. Often these diseases occur in epidemics which may be triggered by changes in climatic conditions that could increase transmission rates (WHO, 2009).

In Akure, Ondo State, Nigeria, it was observed that there was a significant monthly variation in measles occurrence and weather variables. Temperature had the same general pattern of variation with measles, decreasing from January to August and increasing thereafter to December; this suggests that the rate of change in temperature was found to be directly proportional to the occurrence of measles in the state. The pattern was reversed when observed with relationship to relative humidity and rainfall, suggesting a negative relationship (Akinbobola and Omotosho, 2010).

In 2008, an estimated 164,000 people died of measles worldwide; majority of which were children under 5 years of age. About 1-5% of children with measles die from complications arising from the disease. In refugee settings, the death rate from measles usually gets as high as 30%. Targeted vaccination campaigns have had a major impact on minimizing the rate of

deaths caused by measles. From 2000 - 2008, nearly 700 million children aged 9 months to 14 years who live in high risk countries were vaccinated against the disease. Global measles deaths decreased by 78% during this period. (WHO, 2009).

Measles have been devastating in not only developing countries but developed as well. A study in the UK has shown that although various genetic strains of measles are known to exist, recovery from infection from one strain leads to a life time immunity to re-infection from all others. This combination of long lasting immunity and a high infection rate results in measles being a predominantly childhood disease and intimately connects the infection dynamics and the birth process. The short timescale of this progression to immunity leads to relatively violent epidemic cycles, followed by deep troughs in measles abundance. From studies, it has been discovered that the dynamics of the transmission rate of the disease are strongly influenced by seasonal variations (forcing). In the best documented case in the UK, forcing depends on the seasonal aggregation of children in primary schools, thus the pattern of mixing with age is also potentially important, though much can be learned from simple non-age-structured models. In the cases of the less developed countries, more general seasonal drivers other than schooling, such as the agricultural cycle and associated droughts and famine, are likely to play a more significant role (Conlan and Grenfell, 2007).

Immunization from measles is effective, and has resulted in significant reductions in deaths and complications (from the disease) in many parts of the world. Sadly, a large percentage of children in Africa never receive their first measles vaccine dose in time for immunity to take hold. The cost of protecting a child against measles is less than one Dollar, and when correctly administered at 9 months of age, the measles vaccine offers life-long protection to approximately 85% of those vaccinated. (WHO, 2009)

As seen in (Caspi, 2015), three different big data applications were used by three institutions in order to try curb the outbreak of measles. The University Of Pittsburg Graduate School Of Public Health developed an epidemic simulator called FRED Measles Epidemic Simulator. In the simulator, its users can see on a map of any major metropolitan area how a case of measles could become a major outbreak or how it can be quickly quashed, depending on the tare of vaccination in the community. It is available for the public to use freely. According to the institution the simulator is expected in the nearest future to allow users to adjust rate of vaccination in different places in order to see a projected impact of measles

outbreak. More information about the simulator can be seen in (FRED Web, 2019).

The second project is the Bio surveillance Gateway of Los Alamos National Laboratory. This Bio surveillance gateway was a website created for health professionals when there was a measles outbreak. The major aim of the website is to provide support for global disease surveillance providing important tools for health professionals around the globe to access from a single point of reference. The gateway was mainly created for health professionals all over the world who are in need of rapid reference materials so they will know of any potential outbreak of measles and other diseases; this will enable them to try to improve the response times to these outbreaks. Examples of tools include sequence databases, apps for contextualizing disease outbreaks based upon historical data, and advanced bioinformatics software packages, the laboratory announced. Future updates will provide additional tools and data for the technical audience.

The third tool is Text and Media Analytics provided by SAS Health Care Life Sciences. SAS is a business intelligence vendor; they provide a text and social media analytics tool to aid health organizations in tracking conversions involving measles so that they can know when and where there is an outbreak, how it's spreading and how fast it's spreading. With this tool it is possible to spot outbreaks or an epidemic even before the doctors are aware. Also data from the tool can be mined and then used to predict outbreaks of not only measles but other infectious diseases. The tool gathers information about a disease (for example measles) from social media such as tweets, Facebook and blog posts.

In (Idowu, 2013), several data mining models were deployed for predicting five immunize-able diseases that affect children between ages 0-5 years. In their works, Artificial Neural Network, Decision Tree and Naïve Bayes classifier were evaluated for the prediction. The result obtained showed that the use of these models enhanced the effectiveness of routine immunization in Nigeria.

In (Du et al., 2018) a study has been carried out in order to understand the public's perception when there was an outbreak of Measles. The major data that was used for the study were obtained from Twitter and were analyzed using Convolutional Neural Network Models (CNN) and Support Vector Machines (SVM). The work done in this paper is similar to one of the works that was done in (FRED Web, 2019). CNN and SVM were applied to the obtained data and were compared using some metrics. Some of the findings from the study will be discussed. In the study, three major class

distributions were used to classify the perception of people in the times when there were Measles outbreak; these are: discussion themes (resources, personal experience, personal interest, question and other) emotions expressed (Humor or sarcasm, Positive emotion, Anger, Concern, Not applicable) and attitude towards vaccination (Pro, Against, Not applicable). More information about these class distributions are given in the paper.

Using these class distributions the effect of the outbreak of Measles on the human perception was weighed by using CNN and SVM. A total of 1151 tweets were analyzed and it was found out that there was a very high unbalance when comparing the class distributions. With all the vast amount of data analyzed and conclusions reached from the result, it can be seen that the use of Machine Learning is a very important in the healthcare sector and will go a long way in the search for a proper way to predict diseases like Measles and Meningitis.

In (Nguyen, 2016) a research has been carried out by evaluating statistical and Machine learning methods to carry out predictions on the risk of child mortality in Uganda. These child mortality cases are usually caused by dangerous and communicable diseases such as Meningitis and Measles. In this research, data was collected from hospitals and then Machine Learning Methods (such as Random Forests and Logistic Regression) were applied to predict the occurrence of these deadly diseases. The performance of these methods in prediction of these diseases were measured and compared. During the extraction of data, two major statistics were considered – percentage of cases with diagnosis at admission & discharge and number of cases & deaths with diagnosis at discharge and the percentage of deaths. In the first statistic it can be seen that Meningitis and Measles had mere 0.72% and 0.85% respectively. This shows that almost every case at the stage of admission and (even) discharge have not been diagnosed. Even worse is the second statistic; the top disease with the highest percentage of death cases (out of number of cases at admission and discharge) at 18.68% is Meningitis! Measles is down the ladder on this one with 2.24%. With the statistics presented above it is highly of importance for (a) method(s) to be developed for prediction of Meningitis and Measles. In (Opio, 2011) a study was carried out in Uganda to determine how the seasons affect the rate of outbreak of dangerous diseases like Measles. In order to achieve results, clinical data from specific trusted health offices and meteorological data from the Meteorological department of Ugandan Ministry of Water and Environment were obtained and used. In order to work on the obtained data statistical and Machine Learning methods were

used. Linear Regression was employed to test the obtained data. At the end of the research it was discovered that an outbreak of Measles is expected biannually. Also discovered from the research is the season of high and low occurrence of the disease, it is usually high in rainy seasons and low in the dry seasons. Importantly, the researcher was able to ascertain that from the results a model could be developed to estimate the incidences of Measles and also to predict the disease.

RESEARCH METHODOLOGY

Research Framework

The methodology proposed in this research consists of several stages that need to be performed to accomplish the objectives stated below. The discussions are shown in the following sections starting from the problem formulation to the model design and development. This is followed by model testing, performance evaluation and validation. The objectives are :-

- (i) Collect data from the study area on measles and meningitis
- (ii) Design a Non-linear Autoregressive Recurrent Neural Network (NARX NN) with Exogenous Input.
- (iii) Implement the designed model in (ii) with data collected in (i)
- (iv) Develop Random Forest (RF) and Linear Regression (LR) models and apply
- (v) Implement the developed models in (iv) with the data collected in (i)
- (vi) Compare the performance of models implemented in (iii) and (v) using standard metrics like correlation coefficient and root mean squared error.

NARX Model Design

Prediction can be described as a dynamic filtering process in which future values of an event $Y(t)$ is predicted using past values $X(t)$ of one or more-time series of that event. Different techniques tackle prediction differently, but in general, a prediction model or machine learning algorithm learns a target function (f) which best maps input variables X to an output variable Y represented mathematically as:

$$Y=f(X) \quad (3.1)$$

$$Y(t)=f(X(t-1),Y(t-1)) \quad (3.2)$$

$$Y(t)=f(Y(t-1)) \quad (3.3)$$

Equation (3.1) is the general prediction model and used to represent most prediction models like linear regression models and random forest models. Equation (3.2) represents the NARX model where $(t-1)$ represents the

past values of variable X and $Y(t-1)$ represents the past values of series Y. In the event that the past values of variable X are not available, the Non-linear Autoregressive (NARX) prediction model shown in Equation (3.3) can be used. In this research work, the date represents variable X while the number of measles case is Y.

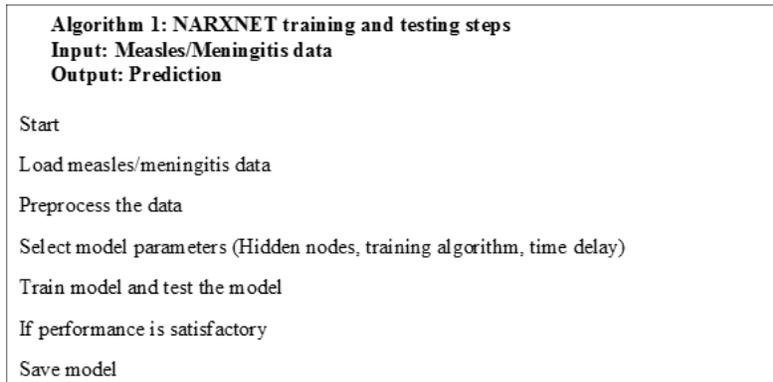
The type of the neural network to be used for the prediction of the diseases is the first design decision that was taken. The NARX model proposed is a 1*72 inputs representing 72 time steps with one elements and 1*72 output representing dynamic data of 72 time steps of 1 element because of the nature of the dataset which comprise of Date and Disease occurrence number. Other parameters and their values set for this model are shown in Table 3.1.

Table 3.1: NARX parameter settings

S/N	PARAMETER	VALUE
1	Number of time step	1
2	Number of delays (d)	2-10
3	Number of hidden nodes	10-50
4	Training Algorithm	(LM, SCG, BR)

Application of NARX for Measles Prediction

The designed NARX model will be trained and tested using the preprocessed datasets already described. The steps for training and testing are



described in the algorithm in Figure 3.1

Figure 3.1: NARX training steps

Development of RF Model

Random forest is one of the most well-known ensemble algorithms that uses decision tree as base classifier. The construction a random forest conforms to the general process of building an ensemble, which consists of three main following phases.

1. Gaining ensemble diversity - Random forest algorithm gains ensemble diversity by manipulating training sets. A list of learning sets is created using the bootstrap sampling method.

2. Constructing base classifier - Random forest employs the same inducer, which is random tree, on different training sets generated in the previous step to build base classifiers. In detail, at each node, a small group of input attributes is selected randomly. The size of the group can be predefined by users, but usually it is chosen as the greatest integer that is not greater than $\log_2 M + 1$, where M is the number of input attributes. Next, the best attribute or the best split point would be selected to split on. All those trees are not pruned.

3. Combining base classifiers - The Majority voting method is utilized in the Random forest algorithm.

Breiman (2001), the "father" of Random Forest, defined it as follows. A random forest is a classifier consisting of a collection of tree-structured classifiers $\{h(x, \theta_k), k = 1, \dots\}$ where the $\{\theta_k\}$; are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x . In other words, building a random forest comprises the task of generating random vectors to grow an ensemble of trees and letting those trees vote for the most popular class. Tran, (2015).

Application of RF Model for Measles Prediction

The steps for applying RF model for measles are as follows:

Algorithm 2: Random Forest Training and testing steps

Input: Measles data

Output: prediction

- i. Start
- ii. For each measles or meningitis data;
 - a. Create 2/3 random samples of measles or meningitis data using bootstrap aggregating

- b. Create Out of Bag (OOB) samples from the remaining 1/3 samples
- iii. Train RF model on the newly created 2/3 data in (ii(a))
- iv. Evaluate the prediction error from OOB samples in (ii(b))
- v. If Stop condition is not met?
- vi. Repeat step (ii) to (iv)
- vii. Compute average prediction
- viii. Stop.

Figure 3.2: RF training steps

Results and Discussion

NARX NN Measles Prediction Result

Figure 4.1 shows the NARX-NN training performance convergence curve. This curve shows the convergence point of the training process from the validated Mean Squared Error (MSE) results which is 119.9372. At this point, the best training and testing results in terms of minimum MSE was obtained. Figure 4.2 shows the time-series response of the NARX-NN training, testing and validation performance. The plot shows the error difference between training outputs and training targets (in blue), validation outputs and targets (in green) and testing outputs and targets (in red). The result shows a minimum error (<10) between the outputs and targets for most of the samples which indicates a better result.

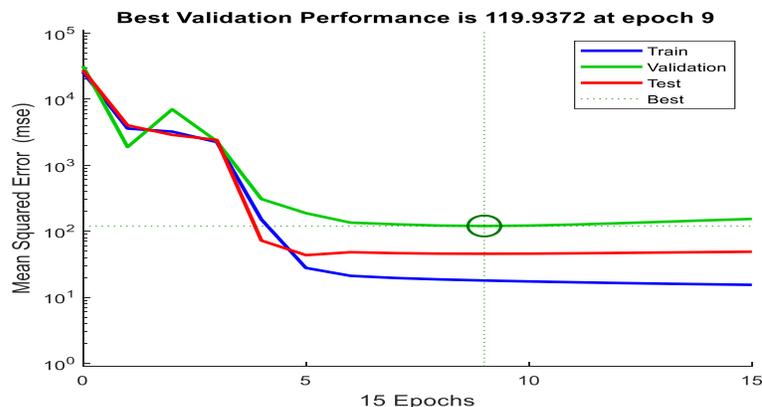


Figure 4.1: NARX-NN Training Performance

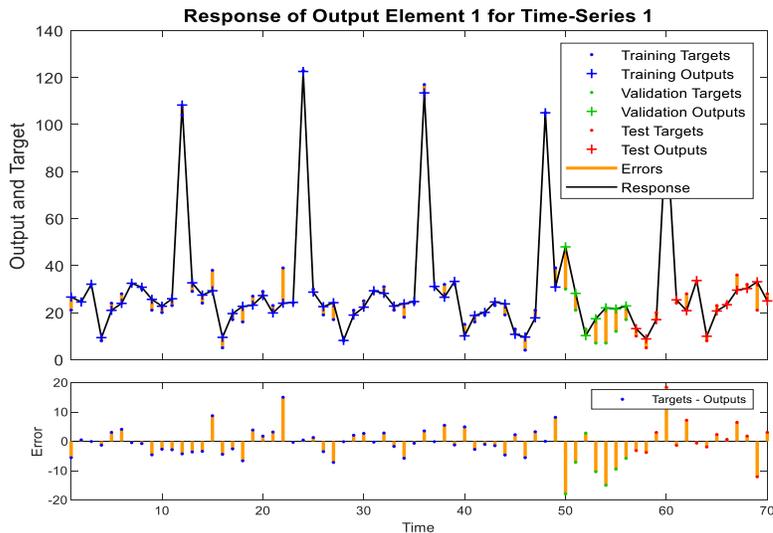


Figure 4.2: NARX-NN Time-Series Response

Figure 4.3 shows the regression plot of the NARX-NN training stage. The plot shows the regression between training output and training targets. A regression value (R) of 1 means the outputs correlate with the target and R=0, means no correlation between output and target. For Measles prediction, the NARX-NN training gives a regression of 0.98673 which indicates the closeness between the output and targets of the training sets.

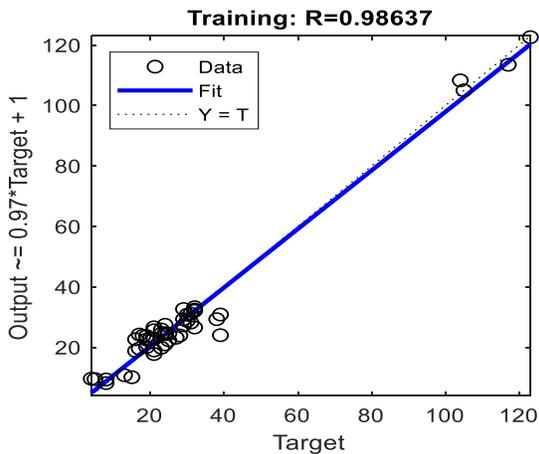


Figure 4.3: NARX-NN Training Regression Plot

Similarly, Figure 4.4 shows the regression plot of the NARX-NN testing stage. The plot shows the regression between testing output and testing targets. For Measles prediction, the NARX-NN testing results gives a regression of 0.98189 which indicates the closeness between the output and targets of the training sets. This results indicate the suitability of the model to be used for prediction of Measles.

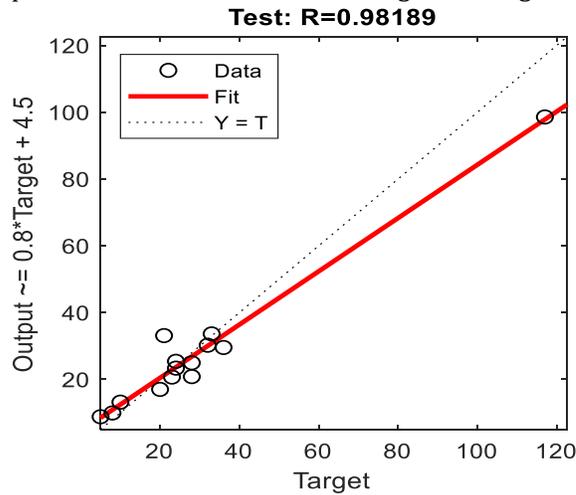


Figure 4.4: NARX-NN Testing Regression Plot

Table 4.1 shows the summary of results obtained by the NARX model for Measles prediction. The root mean squared error (RMSE) obtained is 6.7483 and R of 0.98189.

Table 4.1: NARX-NN Testing Result

MODEL	R	RMSE
NARX-NN	0.98189	6.7483

Random Forest Measles Prediction Result

In this section, the results obtained for the prediction of Measles using RF model is presented. Figure 4.5 shows Random Forest Regression tree and it indicate how rules are fired for classification.

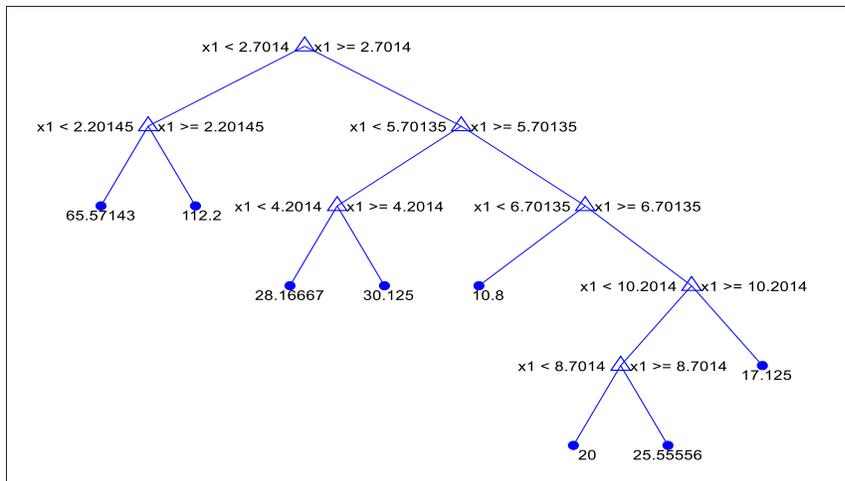


Figure 4.5: Random Forest Regression Tree View

Figure 4.6 shows the regression plot of the RF model testing stage. The plot shows the regression between testing output and testing targets. For Measles prediction, the RF testing results gives a regression of 0.8534 which indicates the closeness between the output and targets of the training sets. This results indicate the suitability of the model to be used for prediction of Measles.

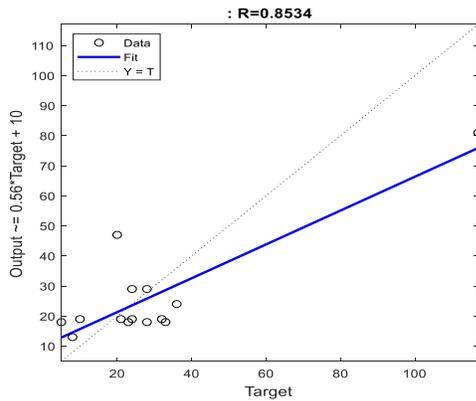


Figure 4.6: RF Test Regression Plot

Table 4.2 shows the summary of results obtained by the RF model for Measles prediction. It states the results of the root mean squared error (RMSE) and R values obtained for different Number of Bags used. It could be seen that for 20 number of bags used, 0.85340 of R value and 14.4469 of RMSE value were obtained. Also for 40 number of bags used, 0.84230 of R value and 14.5480 of RMSE value were obtained as so on for 60, 80, and 100 number of bags. However, it could be seen from the table that the smaller the number of bags, the closer the R value is to 1 and the higher the RMSE values.

Table 4.2: RF Measles Prediction Result

Number of Bags	R	RMSE
20	0.85340	14.4463
40	0.84230	14.5480
60	0.84074	14.6824
80	0.83559	15.9523
100	0.81518	15.4018

Linear Regression Measles Prediction Result

Figure 4.7 shows the regression plot of the LR training stage. The plot shows the regression between training output and training targets. A

regression value (R) of 1 means the outputs correlate with the target and $R=0$, means no correlation between output and target. For Measles prediction, the LR training gives a regression of 0.42673 which indicates the closeness between the output and targets of the training sets. This results indicate non- suitability of the model to be used for prediction of Measles because the value of R is small compared with the values gotten while NARX and RF models were being used.

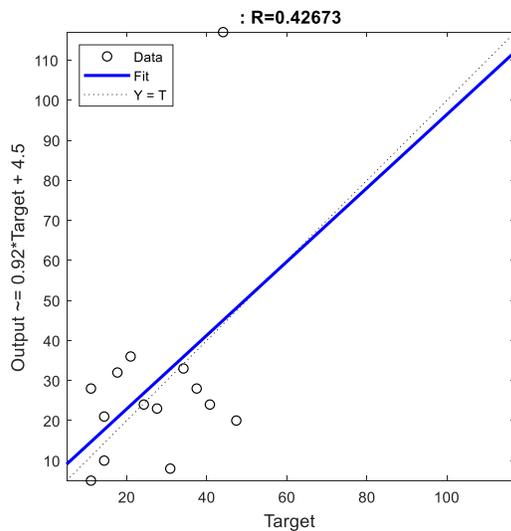


Figure 4.7: LR Test Regression Plot

NARX_NN, RF and LR Measles Prediction Performance Comparison

Table 4.3 shows the summary of results for the comparison of the three models for predicting measles. Comparing the three predicting models, the NARX NN model proved to be better than the RF and LR models as it has a smaller mean squared error of 6.7483 as compared with the mean squared error of the RF which is 14.4463 and LR which is 23.6065. This means that, predicting Measles disease with NARX NN will give closer Regression (R) values than predicting with RF and LR models, since the argument is the closeness of R value to 1, the better the performance.

Table 4.3: NARX-NN and RF Measles Prediction Performance

MODELS	R	RMSE
NARX-NN	0.98189	6.7483
RF	0.85340	14.4463
LR	0.42673	23.6065

Conclusion and Recommendations

NARX NN model was used for predicting measles disease and When compared with RF and LR models, the NARX NN predicting tool proved to be more accurate than the RF and LR models as it had a smaller root mean squared error of 6.7483 as compared to the root mean squared error of the RF which was 14.4463 and that of LR which is 23.6065. The NARX NN model performed better when used in predicting both Measles diseases than RF and LR. However, LR model is not an ideal tool for predicting data with small size as it was not suitable for diseases predicted in the research work. The NARX NN model can be used to predict diseases in the various Health Care Centers and also agricultural produce and can be used to guide the clinicians and other researchers to take the most appropriate decision.

References

- Adebayo, P. I., Bernard, I. A., Olajide, O. A., Aderonke, A. K., & Adekemi, F. O., (2013). Data mining techniques for predicting immunize-able diseases: Nigeria as a case study. *international journal of applied information systems(IJAIS)*, (5)7, 5-15. Retrieved from <http://www.ijais.org>
- Akinbobola D. and Omotosho, J. B. (2010). Meteorological Factors and measles occurrence in Akure, Ondo state, Nigeria, *Journal of Meteorology and related Sciences* 4, 35-47.
- Aličković, E., & Subasi, A. (2011). Data Mining Techniques for Medical Data Classification. *The International Arab Conference on Information Technology (ACIT)*.
- Baker, R., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *JEDM | Journal of Educational Data*

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE
(VOL. 10 NO.1) JUNE, 2019 EDITIONS

- Mining*, 1(1), 3-17. Retrieved from <https://jedm.educationaldatamining.org/index.php/JEDM/article/view/8>
- Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.
- Caspi, H. (2015, March 18). *3 Big Data Initiatives that are Targeting the Measles Outbreak*. Retrieved from Predictive Analysis Times - Machine learning & data science news: <https://www.predictiveanalyticsworld.com/patimes/3-big-data-initiatives-that-are-targeting-the-measles-outbreak318/5092/>
- Clark, K. a. (2007). Kumar and Clark's clinical medicine. In K. a. Clark, *Kumar and Clark's clinical medicine*. England: Saunders.
- Coates, A., Karpathy, A., & Ng, A. Y. (2012). Emergence of object-selective features in unsupervised feature learning. In *Advances in Neural Information Processing Systems* (pp. 2681-2689).
- Conlan A.J.K.and B.T.Grenfell, 2007; Seasonality and the persistence and invasion of Measles, PMC, PubMed.
- FRED Web*. (2019). Retrieved from FRED Web: <http://fred.publichealth.pitt.edu/measles/>
- El-Halees, Alaa M. (2007). Arabic Text Classification Using Maximum Entropy. IUG Journal for Natural and Engineering Studies, Volume: 15, Number: 1, غرة - ية سلام الإ . <http://hdl.handle.net/20.500.12358/23860>
- Gayathri, P. and Jaisankar, N. (2015). "Medical Data Mining Applications and its Uses", International Journal of Engineering Technology Science and Research IJETSR, Volume 2, Issue 6, pp. 8 -14.
- Han, J. and Kamber, M. (2006). "Data mining: concepts and techniques", 2nd edition. The Morgan Kaufmann Series.
- Ji, X. (2014). Social Data Integration and Analytics for Health. Newark, NJ, 07032, USA.
- Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning* (pp. 137-142). Springer, Berlin, Heidelberg.

**INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE
(VOL. 10 NO.1) JUNE, 2019 EDITIONS**

- Kale, T. D. (2015). Application of Data Mining Techniques to Discover Cause of Under- Five Children Admission to Pediatric Ward: The Case of Nigist Eleni Mohammed Memorial Zonal Hospital. *Journal of Health & Medical Informatics*, 6(178). doi:10.4172/2157-7420.1000178
- Kautish, S., & Ahmed, R. K. (2016). A Comprehensive Review of Current and Future Applications of Data Mining in Medicine & Healthcare. *International Journal of Engineering Trends and Technology (IJETT)*, 38(2) 60-63. doi: 10.14445/22315381/IJETT-V38P212
- Koh, H. C. and Tan, G. (2005). "Data Mining Application in Healthcare", *Journal of Healthcare Information Management*, spring, vol. 19, issue. 2, pp. 64-72.
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160, 3-24.
- Mantzaris, D., Anastassopoulos, G., & Lymperopoulos, D. (2008). Medical Disease Prediction Using Artificial Neural Networks. doi:10.1109/BIBE.2008.4696782
- Nigam, K., Lafferty, J., & McCallum, A. (1999, August). Using maximum entropy for text classification. In *IJCAI-99 workshop on machine learning for information filtering* (Vol. 1, No. 1, pp. 61-67).
- Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques*. Springer Science & Business Media.
- Patil, Sheetal L. (2015). "Survey of Data Mining Techniques in Healthcare", *International Research Journal of Innovative Engineering*, Volume1, Issue 9, pp. 1-3.
- Que T., (2015). Improving Random Forest Algorithm through Automatic Programming. *Master Thesis in Computer Science*, 27, Retrieved from www.hiof.no
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.
- Rish, I. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46).

**INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE
(VOL. 10 NO.1) JUNE, 2019 EDITIONS**

- Rokach, L., & Maimon, O. Z. (2008). Data mining with decision trees: theory and applications (Vol. 69). World scientific.
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380-1400.
- Singh, R., Singh, R., & Bhatia, A. (2018). Sentiment analysis using machine learning techniques to predict outbreaks and epidemics. *International Journal of Advanced Science and Research*, 3(2), 19-24. Retrieved from www.allsciencejournal.com
- Srinivas, K., Rani, B. K., & Govrdhan, A. (2010). Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. (*IJCSE*) *International Journal on Computer Science and Engineering*, 2(0975-3397), 250-255.
- Sumathi, S., Mohanapriya, S., Nagasandhiyalakshmi, B. & Shanmugapriya, N., (2016). Prediction of outbreak of heart disease using Text Mining. *National Conference On Advance Trends In Information Computing Sciences (NCATICS '16)*, 52(245), 1070-1077.
- Thenmozhi K., Deepika, P.(2014). "Heart Disease Prediction Using Classification with Different Decision Tree Techniques", *International Journal of Engineering Research and General Science*, Vol. 2, Issue 6, pp. 6-11.
- Vashi, I. R., Mishra, S., & Andhariya, S. (2017). A Comparative Study of Classification Algorithms for Diseases Prediction in medical domain. *Asian Journal of Convergence in Technology*, 3(3).
- 56
- Vapnik, Vladimir N. (1995). *The Nature of Statistical Learning Theory*. Springer, New York.
- Werts, N. and Adya, M. (2000). "Data Mining in Healthcare: Issues and a Research Agenda", the proceedings of the conference AMCIS 2000 Proceedings, Association for Information Systems (AIS), pp. 94-97
- Zhang, H. (2004). The optimality of naive Bayes. *AA*, 1(2), 3.