www.idosr.org Olum and Okeke
©IDOSR PUBLICATIONS
International Digital Organization for Scientific Research ISSN: 2550-7931
IDOSR JOURNAL OF APPLIED SCIENCES 8(1) 1-12, 2023.

Prediction of Covid-19 New Cases and New Deaths in Nigeria Using Polynomial Regression Model

[1]Olum Ekene J. and [2]O.C. Okeke
[1,2]Department of Computer Science, Chukwuemeka Odumegwu Ojukwu University
[1]olumekene@gmail.com; [2]ogoookeke@yahoo.com

ABSTRACT

The Coronavirus disease 2019 (COVID-2019) is a global pandemic and Nigeria is not left out in being affected. This study is focused on the analysis prediction of the spread of Covid-19 in Nigeria, applying statistical model (Polynomial Regression) and available data from the NCDC. The study is an insight into the spread of Covid-19 in Nigeria in order to establish a suitable prediction model, which can be applied as a decision-supportive tool for assigning health interventions and mitigating the spread of the Covid-19 infection. The polynomial regression models are prominent tools in predicting the impact of certain factors on COVID-19 outbreak and taking the necessary measures to respond to this crisis. The objective of this study is to implement a predictive model in forecasting the near future number of new death and positive COVID-19 cases expected in Nigeria following the present trend. A polynomial regression model prediction on the epidemiological data obtained from Nigerian Centre for Disease Control to predict the epidemiological trend of the prevalence and incidence of COVID-2019. In this study, we examined the inadequacies involved in the manual method of predicting new death and COVID-19 case to enable government to make decision in reducing the death and spread of COVID-19 in Nigeria. The methodology we adopted in this study is the Object-Oriented Design Methodology (OODM) which combines data and processes (methods) into single entities called objects while Microsoft Visual Basic is our programming language of implementation. The new system will be used in generation of reports and results to create, process, and validate a model that can be used to forecast future outcomes of new death and COVID-19 case. The system was implemented using VB.net programming language and SQLite because it is not a server side application, which needs a standard database management system.

Keywords: Prediction, COVID-19, Coronavirus, Pandemic, Morbidity, Polynomial regression, Prominent, Predictors, Nigerian Centre for Disease Control (NCDC).

INTRODUCTION

The coronavirus outbreak is the most notable world crisis since the Second World War. The pandemic that originated from Wuhan, China in late 2019 has affected all the nations of the world and triggered a global economic crisis whose impact will be felt for years to come [1,2]. COVID-19 has been spreading rapidly globally, with a considerable impact on global morbidity, mortality and healthcare utilization. This viral disease is caused by the Coronaviruses (CoV) which belong to the genus 'coronavirus' of the Coronaviridae family [3,4]. All CoVs are pleomorphic RNA viruses characterized by crown-shape (the name "coronavirus" is derived from the Greek κορώνα, meaning crown.) peplomers with 80-160 nM in size and the genome of CoV contains a linear, single-stranded RNA molecule of positive (mRNA) polarity and about 28-32Kb in length [5,6]. Coronaviruses possess the largest genomes among all identified RNA viruses, and the large genome gives the virus extra plasticity in accommodating and modifying genes [7,8,9].

The COVID-19 disease is current ravaging the world and Nigeria is not left out in this global pandemic. Since the first case in late February, 2020 in the country, there has been an increase rise in the number of cases declared by the agency responsible for tracking, monitoring and instituting management guideline in Nigeria [10, 11, 12]. The first COVID-19 confirmed case in Nigeria was reported on 27th February 2020, when an Italian citizen in Lagos tested positive for the virus [13] The second case was recorded on the 9th March 2020 in Ewekoro, Ogun State, a Nigerian citizen who had contact with the Italian citizen [14]. Within the first month, the confirmed cases were around 70 but drastically increased to number almost 1,350 cases before the end of the second month. The discharged cases increased from 3 to about 250 in the first two months. The number of recorded deaths increased from 1 to 40. Prediction model is the process of using known results to create, process, and validate a model that can be used to forecast future outcomes. Prediction model uses statistics to predict outcomes [15]. Most often the event one wants to predict is in the future, but predictive modelling can be applied to any type of unknown event, regardless of when it occurred. Predictive analytics uses predictors or known features to create predictive models that will be used in obtaining an output. A predictive model is able to learn how different points of data connect with each other. Two of the most widely used predictive modeling techniques are regression and neural networks. In the field of statistics, regression refers to a linear relationship between the input and output variables. A predictive model with a linear function requires one predictor or feature in order to predict the output or outcome. For my model, I chose the Polynomial Regression Model (PR) model. In this model, the future value of a variable is assumed to be a linear function of several past observations and random errors. In statistics, polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted E(y |x). Although polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear, in the sense that the regression function E(y | x) is linear in the unknown parameters that are estimated from the data. It is also called the special case of Multiple Linear Regression in Machine Learning. It is a linear model with some modification in order to increase the accuracy. For this reason, polynomial regression is considered to be a special case of multiple linear regression.

Statement of Problem

1. COVID-19 is a new pandemic triggered by extreme acute respiratory coronavirus syndrome 2 (SARS-CoV-2) and spreads rapidly from person to person. The coronavirus outbreak is the most notable world crisis since the Second World War.
2. The pandemic that originated from Wuhan, China in late 2019 has affected all the nations of the world and triggered a global economic crisis whose impact will be felt for years to come. Different models have been used in recent studies to predict incidence, prevalence and mortality rate of COVID-19.
3. The problem is that there is no cure for infection with the coronavirus behind the COVID-19 pandemic. All we have to do is to learn how to prevent it from spreading from one person to another. With the help of prediction model, we can know how far and how good the effort of stopping the spread of the disease COVID-19.

Aim and Objectives of the Study

The aim of this study is to build a new system for predicting the new death and COVID-19 death cases using polynomial regression model.
The objectives are as follows:

i. To use polynomial regression model algorithm to predict or forecast some probabilistic attributes of new death cases and COVID-19 cases.

ii. To develop a predictive system that will reduce the risk of death and COVID-19 death cases in Nigeria.
iii. To design an advance desktop base predictive system that will help government to improve operation in reducing new death and COVID-19 death cases using polynomial regression.

Literature Review

Prediction research, which aims to predict future events or outcomes based on patterns within a set of variables, has become increasingly popular in medical research. Accurate predictive models can inform patients and physicians about the future course of an illness or the risk of developing an illness and thereby help guide decisions on screening and/or treatment. For example, predictive models have been developed in gastroenterology to predict the risk of disease flares for inflammatory bowel disease and risk of hepatocellular carcinoma among patients with cirrhosis [16, 17, 18]. There are several important differences between traditional explanatory research and prediction research. Explanatory research typically applies statistical methods to test causal hypotheses using a priori theoretical constructs (e.g., hepatocellular carcinoma surveillance underutilization is related to provider-level factors4). In contrast, predictive research applies statistical methods and/or data mining techniques, without preconceived theoretical constructs, to predict future outcomes (e.g., predicting the risk of hospital readmission. Although predictive models may be used to provide insight into causality of pathophysiology of the outcome, causality is neither a primary aim nor a requirement for variable inclusion. Noncausal predictive factors may be surrogates for other drivers of disease, with tumor markers as predictors of cancer progression or recurrence being the most common example. Unfortunately, a poor understanding of the differences in methodology between explanatory and predictive research has led to a wide variation in the methodologic quality of prediction research. The aim of this primer is to describe basic methods for conducting prediction research, which can be divided into three main steps: developing a predictive model, independently validating its performance, and prospectively studying its clinical impact [19, 20].

Types of Predictive Models

Although prediction research in medicine has traditionally used a Bayesian framework approach, with statistical techniques such as regression models, data mining techniques such as machine learning algorithms are a form of artificial intelligence that are being used with increasing frequency. Machine learning has been previously used to predict behavior or outcomes in business, such as identifying consumer preferences for products based on prior purchasing history. A number of different techniques to develop predictive algorithms exist, using a variety of prediction analytic tools/software and have been described in extensive detail elsewhere [21]. Some examples include neural networks, support vector machines and decision trees. Decision trees, for example, use techniques such as classification and regression trees, boosting and random forest to predict various outcomes. The analysis can be conducted using free software environments such as ''R''10 as well as vendor applications [22]. Machine learning algorithms, such as random-forest approaches, have several advantages over traditional explanatory statistical modeling, such as lack of a predefined hypothesis, making it less likely to overlook unexpected predictor variables or potential interactions. Approaching a predictive problem without a specific causal hypothesis can be quite effective when many potential predictors are available (increasingly common with electronic health records) and when there are interactions between predictors, which are common in biological and social causative processes. Predictive models using machine learning algorithms may therefore facilitate recognition of clinically important risk and variables in patients with several marginal risk factors that may otherwise not be identified. In fact, many examples of discovery of unexpected predictor variables exist in the machine learning literature [23].

Developing a Predictive Model

The first step in developing a predictive model, when using traditional regression analysis, is selecting relevant candidate predictor variables for possible inclusion in the model; however, there is no consensus for the best strategy to do so [24]. A backward-elimination approach starts with all candidate variables, and hypothesis tests are sequentially applied to determine which variables should be removed from the final model, whereas a full-model approach includes all candidate variables to avoid potential overfitting and selection bias. Previously reported significant predictor variables should typically be included in the final model regardless of their statistical significance but the number of variables included is usually limited by the sample size of the data set. Inappropriate selection of variables is an important and common cause of poor model performance in this situation. As described above, variable selection is less of an issue using machine learning techniques given that they are often not solely based on predefined hypotheses. There are several other important issues related to data management when developing a predictive model, such as dealing with missing data and variable transformation; however, these topics are beyond the scope of this primer and addressed elsewhere [25].

Validating a Predictive Model

For a prediction model to be valuable, it must not only have predictive ability in the derivation cohort but must also perform well in a validation cohort.7,18 A model's performance may differ substantially between derivation and validation cohorts for several reasons including overfitting of the model, missing important predictor variables, inter observer variability of predictors leading to measurement errors, and differences in the patient cohort case mix.18 Therefore, model performance in the derivation cohort may be overly optimistic and is not a guarantee that the model will perform equally well in new patients. For example, external validation of the HALT-C predictive model for hepatocellular carcinoma was recently demonstrated to have a significantly worse performance in an external validation cohort.3 Unfortunately, the majority of published prediction research focuses solely on model derivation, and validation studies are scarce [26].

Validation can be performed using internal or external validation. A common approach to internal validation is to split the data set into two portions a ''training set'' and ''validation set''. If splitting the data set is not possible given the limited available data, measures such as cross validation or bootstrapping can be used for internal validation. Machine learning algorithms, more specifically the random-forest approach, uses an alternative approach called''in-bag'' and ''out-of-bag'' sampling. In a random-forest approach, the initial cohort is divided into two groups''in-bag'' and ''out of-bag'' samples. The in-bag sample is created using random sampling with replacement from the initial cohort, creating a sample equivalent in size to the initial cohort. The out-of-bag sample is composed of the unsampled data from the initial cohort, and typically includes about one-third of the initial cohort. The ''out-of-bag'' cohort can serve as an internal validation cohort for the model derived using the ''in-bag'' sample. However, internal validation nearly always yields optimistic results given that the derivation and validation data sets are very similar (as they are from the same cohort). Although external validation is more difficult as it requires data collected from similar patients in a different setting or a different center, it is always preferred to internal validation [27]. When a validation study shows disappointing results, researchers are often tempted to reject the initial model and to develop a new predictive model using the validation cohort data. For example, there are over 60 published predictive models for breast cancer. This approach neglects the information captured from prior studies and predictive models. There are several methods to update prior predictive models with data from the patients of the validation cohort, but these are unfortunately rarely utilized [28].

Models and Algorithms used in predictive analytics

Thus, predictive models are creating during the predictive modelling process to discover the patterns between dependent variables and explanatory variables and predicting an outcome [9].

Various algorithms are used:

Classification: decisive outcome, it's for predicting the value of decisive variable (class or target) by constructing a model based on one or multi decisive or numerical variables (attributes or predictors).

Clustering (unsupervised learning): assigning observations into clusters, each cluster contains the similar observations and data. This process helps in discovering the unknown relationships in a dataset.

Association rules: to find important associations in the observations, which mean association rules find all item sets that have support greater than the minimum support and then using the large item sets to generate the desired rules that have confidence greater than the minimum confidence. An example of association rules application is market basket analysis which is a modelling technique that can be described simply as if a customer buys a specific set of items he will more or less probably buy another set of items.

Regression: numerical outcome, predicting the value of target (numerical variable) by constructing a model based on one or more predictors (numerical and categorical variables). Actually, there are different families of these algorithms and different ways of measuring the error as show figure 1 below.

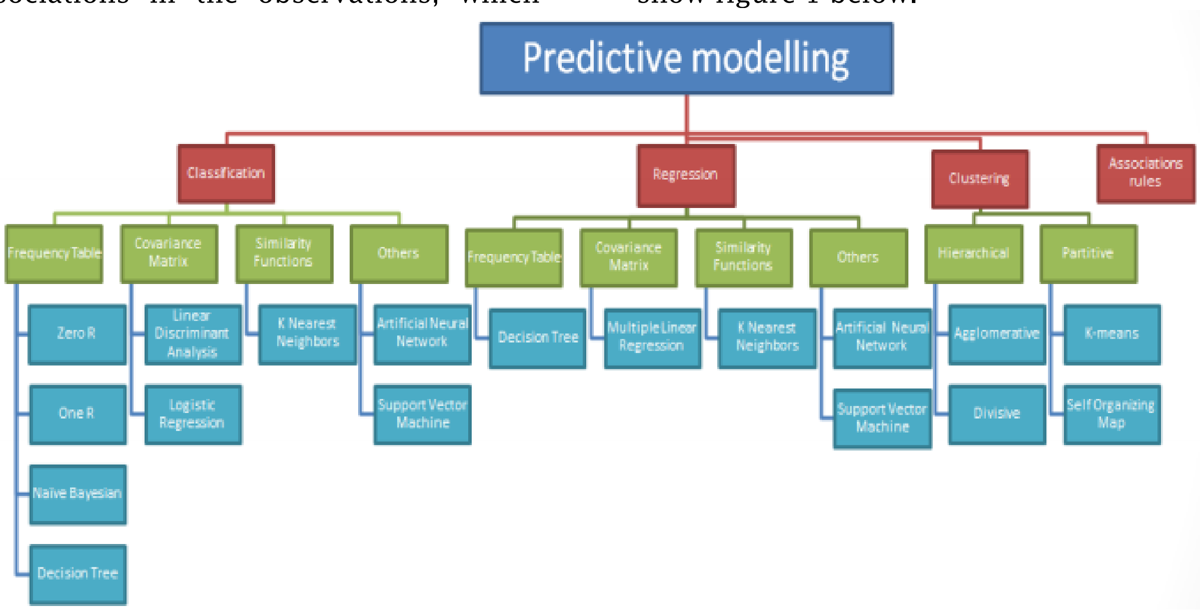

Figure 1: Families of algorithms used in predictive modelling

Review of related works

Predictive analytics in general are used to detect the relationships and patterns in data in order to predict the future by analyzing the past and taking better preventive decisions. Thus, the predictive analytics aim of use differ from one industry to another, for instance a marketer can use the predictive analytics to predict the customers' response to an advertising campaign, or a product seller can use it to predict the movement of product prices, or it can be used to detect trends such as in banks if a manager wish to recognize the most profitable customers, or alert a credit card customer to a probable fraudulent charge. Thus the predictive analytics help in answering many questions such as what will happen if the demand of products decrease? Or if suppliers' prices increase? What is the risk to lose money a new business? [9].

In fact, in the education sector as one of the important sectors the predictive analytics have been used for different purposes and by using different models and tools. For example, a study focus in behavioral analytics in university in order to predict whether students are vulnerable to deviant ideologies which can lead to terrorism [11]. While, another study shows the necessity and benefit from using the predictive analytics in the educational sector which will help the educational institutions to increase the retention of student and enhancing their results and achievements. The researchers focused on the use of classification algorithms especially the decision tree to get better prediction results. The lack in this study is that it's not specific and not tested by real world data [12]. In the same industry, another research aimed by the use of a performance analytic methodology to predict the final students' performance in a specific course during the semester and mark the ones that will fail and have low performance in exams. The researchers used the decision tree algorithm to predict the students' final performance results. Actually, this research will be more beneficial by doing deeper analysis of discovering the students' mistakes in the exams, their learning interactions with the educational system by using different algorithms such as association rule algorithm [13]. Moreover, Hina analyzed the students' data in order to predict the drop out feature of students and discovering the main factors that influence the open sources dropping by students. Thus, to do this analysis the researchers applied feature selection algorithms by using WEKA tool and then classification algorithms. Thus, the results will be more accurate with the use of different algorithms such as association and clustering techniques [17].

Actually, predictive analytics have been used also in business and market in general and in different countries. For instance, one of the studies done on one of the governmental organizations in Singapore described the characteristics of the procurement dataset specifics and its implications on the future purchase problem and solve it by using Markov chains model. In fact, the aim of the study was to cover the greater portion of the future purchasing by predicting the larger number of requesters. To attain this goal some algorithms have been evaluated such as probability distribution analysis, simple random sampling, sequence analysis, and Markov chains. Although, after the test the best approach among the algorithms used was Markov chains but in general the results were unsatisfactory to be practically applied. Thus, for requesters where prediction was possible (accuracy > 0) Markov approach got 0.31 accuracy with 6.64% dataset coverage without clustering; and 0.26 accuracy with 34.27% coverage with clustering. In fact, the integration of clustering reduced the accuracy, but it notably assisted to make predictions for more requesters. This research could get better results with the use of association rules algorithms in order to discover better the relation between the study variables [19]. In India, a study was done in the Indian green coffee supply chain to develop a multi criteria decision support system based or predictive analytics to help stakeholders having better purchases and the ability to take better sales decisions and knowing the requirements of the green coffee supply chain market. Thus, the system must be more dynamic due to rapid market changes [29].

Moreover, in the IT industry a study used predictive analytics to build a model that analyze the factors affecting the chances to loss or win the IT services deals for IT service providers. This study benefits could be to better take decisions for the preparation and allocation of needed resources, and the attributes of sales pursuit may give awareness to what to do to raise the chances to win sales pursuits in the future. The model developed have been applied in two ways, firstly in the beginning of prioritization of the validated deals in the list, thus the model was utilized to give and early ranking list of deals with the chances of winning. While, the second way to apply the model was by the ideal distribution of sales force to following deals in order to increase the sales revenue. Thus, the model be improved by providing online prediction of deal outcome during the sales pursuit process and it can be extended by integrating this predictive model in another optimization model to determine the ideal bidding price [21]. Additionally, predictive analytics have been widely used in manufacture sector. For instance, in one of the research in manufacturing in metal cutting industry aiming to predict power consumption by utilizing big data infrastructure, the researchers created a prototype system by utilizing open platform solutions comprising Hadoop Distributed File System (HDFS), Map Reduce, and a machine-learning tool. The researchers adopt a data-driven analytic modeling approach based on feature vectors which are n-dimensional vectors of numerical or nominal features that classify a machining operation. But the lack in the study was to acquire real data and integrating the analytic systems and the big data infrastructure and the integration of optimization modeling [26]. Furthermore, in a different way of use in a study done in Pakistan the focus was to minimizing the loss of human life from the drone attack by predicting the future attack frequency and the prospective losses and injuries and its adoption by the government. The tool used to build the predictive models was the IBM SPSS, and the selection of the predictive algorithm needed and its parameters was automatic. Actually this study would get better results if the selection of algorithms was manual to get better accuracy [26].

Moreover, predictive analytics can be used in the social media by social media companies and benefit from social media data while, show the importance to extract valuable information from the social media data and use it for the creation of predictive models. Thus, the framework created merge data from multi social media sources for the analysis and it integrate feature selection, similarity metrics and sentiment and trend analysis by using R, WEKA, D3 and JSON [22]. While, in a different sector the predictive analytics have been utilized in transportation, where researchers present a smart public transportation decision support system to predict the times of bus arrival in short and long term. Thus to attain this, the researchers used clustering model to detect the patterns of bus performance, then a real-time vehicle schedule commitment and prediction model to identify the time of bus arrival and irregular operations, then the approach have been empirically validated by utilizing real-world data. Thus the model show reduction error and improvement in predicting delay. Furthermore, in another sector which is the stock market where the researchers created a model to optimize prediction of products and stock market indications. Thus this model allows to set the stock indications future values and trading of financial services which will allow investors to increase significantly their returns on investment and reduce the risk [17]. Finally, one of the important industries that use the predictive analytics is the healthcare sector. For instance, a study in an Australian hospital was designed to develop a framework and a prototype to benefit from the Business Analytics techniques in the context of oncology and cancer care [20].

System Analysis and Methodology

The methodology used in this research is Structured System Analysis and Design Methodology (SSADM). The SSADM consist of three vital techniques which includes; logical data modeling, data flow modeling and entity behavior modeling.

Logical Data Modeling (LDM)

LDM is the process of identifying, modeling and documenting the data requirements of the system being designed.

Data Flow Modeling (DFM)

DFM is the process of identifying, modeling and documenting how data move around in a system. Data flow modeling examines processes, data stores, external entities and data flow.

Entity Behavior Modeling (EBM)

EBM is the process of identifying, modeling and documenting the event that affect each entity and the sequence in which it occurs. There are the various phases/steps that are also implemented in structured system analysis and design methodology. The step involves the following:

i. Problem identification
ii. Feasibility studies
iii. System analysis
iv. Design
v. Coding and testing
vi. Implementation

1. Problem identification: This defines the problem to be solved and this set the direction for the whole project. It also sets the project limits, which defined what part of the system, can be changed by the project. And what parts are outside its control. In this research work the problem identified is that COVID-19 figures in Nigeria are not accurate and standard because of non-availability of tools to extract hidden basic attributes from a large data warehouse to give a 100% accuracy of the prediction. The scope is that data from NCDC is used as input in development of the system.

2. Feasibility study: This proposes one or more conceptual solution to problem set for the project. The conceptual solution gives an idea of what the new system will look like. The researcher would use a decision tree algorithm to predict some basic attributes.

3. System analysis: This phase is a detailed appraisal of the existing system. This appraisal includes finding out how the system works and what it does as well as system problems and solution to the problems. So in the analysis of the present system, it was noticed that all the NCDC prediction in Nigeria was done manually.

4. Coding and testing: The individual program components are written and tested, and user interface developed and tested by programmer. The proposed system is developed using C# language.

5. System design: This phase produces a design for the system. Designers must select equipment needed, structures and detailed documentation on how to use the system.

6. Implementation: During implementation are put into operational used. Usually this means that the new system work in parallel for same time. In this research, the new system has been tested, corrections made and it is implemented.

Analysis of the Proposed System

There is a need to introduce a new system to mitigate the problems inherit with the operations of the old system. The new system will use polynomial regression model to predict some attributes of COVID-19 new cases like States Affected, Lab Confirmed, On Admission, Number Discharged and Number of Deaths. The new system gives map of the predicted output preview using Nigeria National map with dotted signal that will show details of the states with the highest COVID-19 cases. The difficulty with the manual overlay method was that they may be published at different scales or projections. The more layers of maps included in the present analysis and the more complex they become, the more the likelihood of human error entering the analysis and the longer the process takes but not so in the new system. The system will bring in digital technique in the storage and management of data. This will make data movement within department and outside the department fast and effective. The digital format of data will reduce the likely hood of data loss as it moved from official to official; department of a local area network to effect the data movement. This form of storage will also enable the implementation of access control to prevent unauthorized access. This storage format makes it feasible to apply data-mining technology to the analysis of COVID-19 data. The application of data-mining to the COVID-19 analysis would reveal hidden pattern in the population distribution that conventional technique cannot reveal. This could drive more informed decision by government and industry.

Advantages of the proposed system

The new system is fast, reliable, efficient and accurate since all operations are done in a digital format. With polynomial regression model prediction and data-mining technology, hidden patterns are obtained, accurate, standard and actual geographical distribution of COVID-19 are seen with help of GIS map, which help in economic, social and educational development of a Nigeria. This work also showed that since accurate and standard COVID-19 data are predicted; it would help governments, organizations, individuals to foreseen the danger output COVID-19 and plan to prevent the causes before it come to pass.

RESULTS

These are outputs of the workability of the tested system. They give evidence of a fully implemented system when tested with real-time data.

Input form design of the proposed system

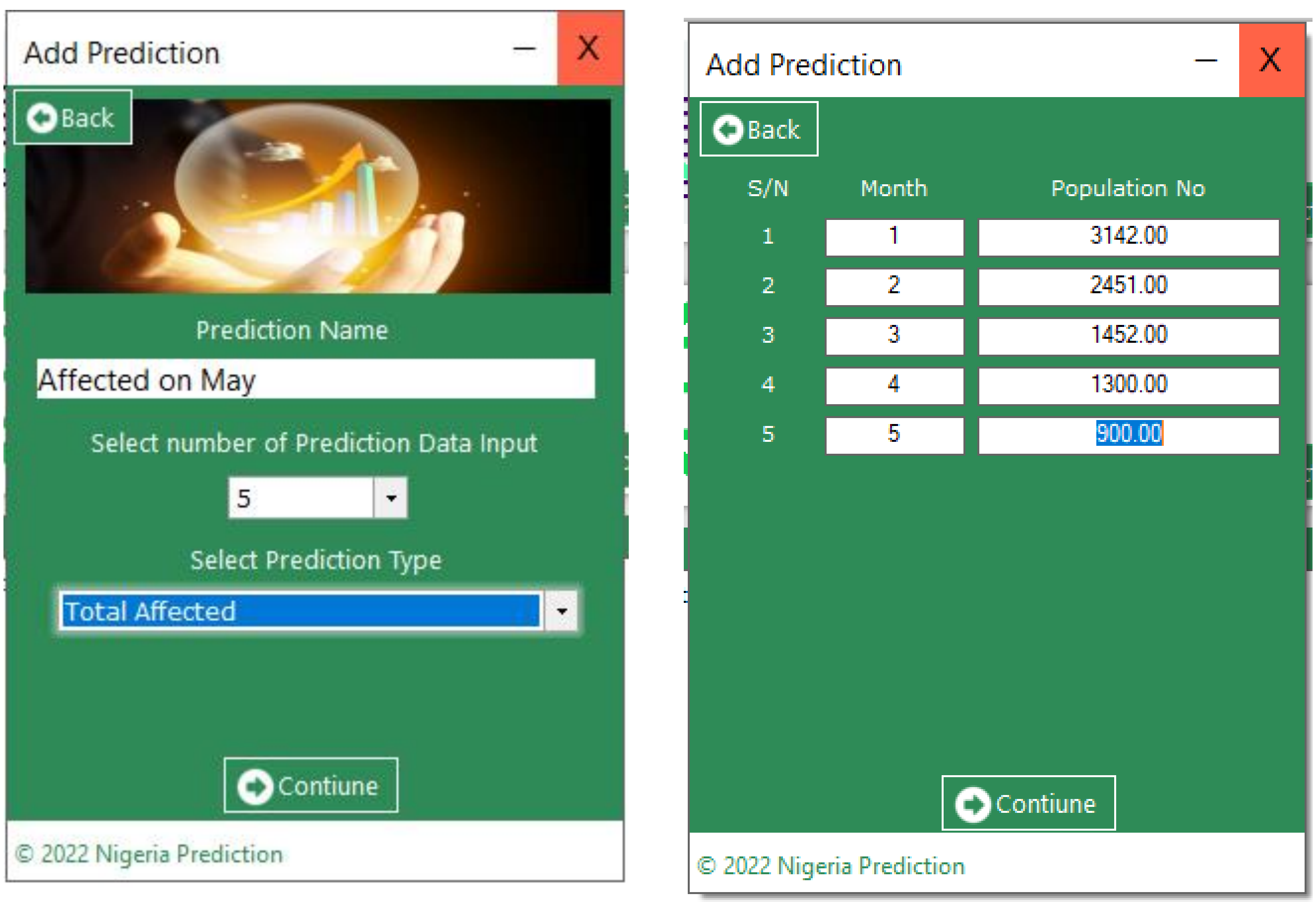

Fig 2 Output form design of the proposed system

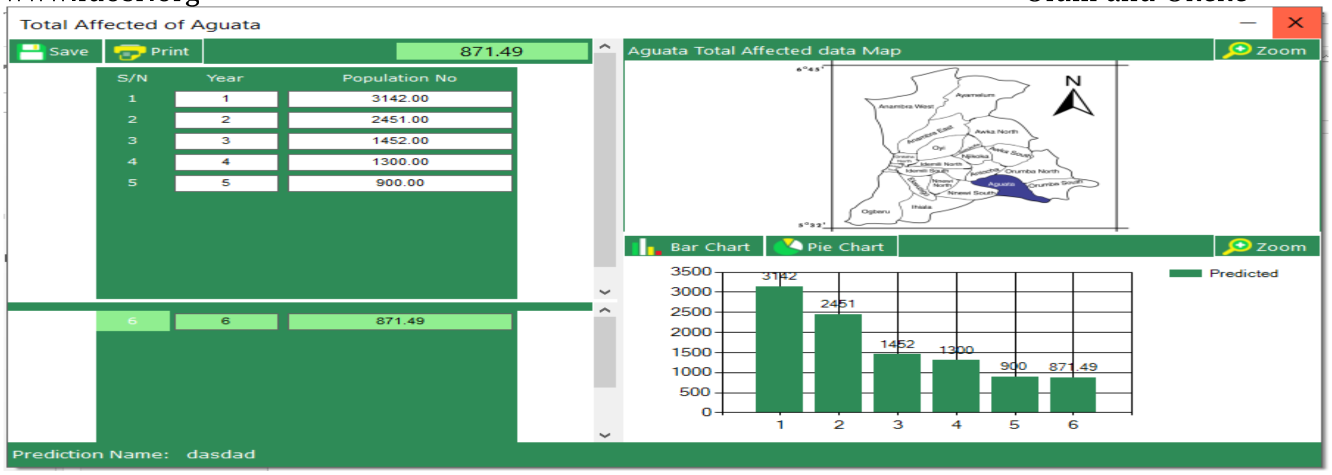

Fig 3: Total Affected of Aguata

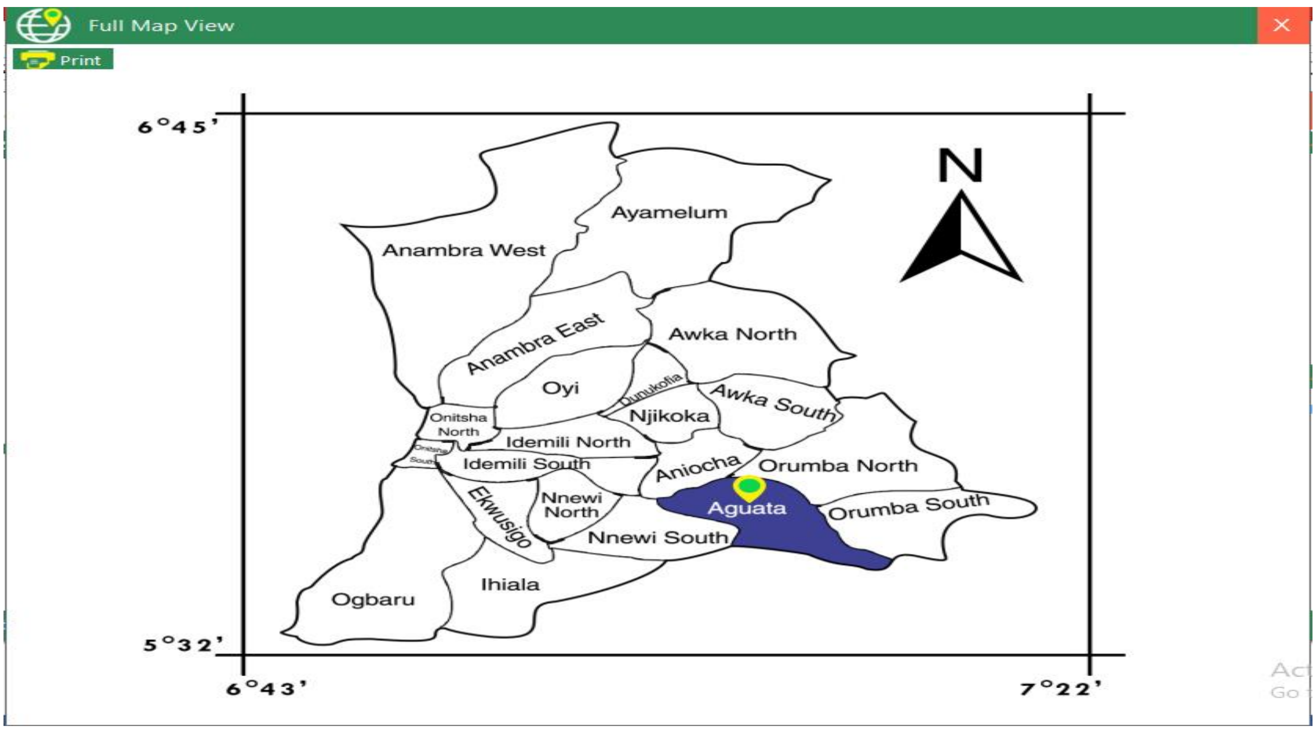

Fig 4: Full Map

DISCUSSION AND CONCLUSION

This dissertation deals with the challenge of creating a model for prediction of COVID-19 new cases and new death in Nigeria. This includes the development of a software interface that will facilitate the prediction of data using polynomial regression model. The map of Nigeria metropolis together with their legend was displayed. Dot density was used to represent distribution of the affected people from (2019 and 2021) in the area. The old system was analyzed and its draw backs were critically pointed out and new system dealt with problems in old system, which helps government to be proactive in decision making.We also note with caution that our data points are still few as the disease is still evolving in the country and with limited testing capabilities, the number of cases declared cannot reflect the actual values of new cases across the country. We know that as the country ramp up her testing strategies and with more data points, our polynomial progression model will continue to improve on the forecast. Statistical methods and the time series models have been adopted in this studies to predict epidemic cases. The polynomial regression model is an essential analytical tool for prediction. In this study, we applied the polynomial regression model to assess the impact COVID-19 new cases and new deaths to confirmed cases in Nigeria.

REFERENCES

1. Adam Westerskia, Rajaraman Kanagasabaia, Jiayu Wongb, and Henry Chang. (2015). *Prediction of Enterprise Purchases using Markov models in Procurement Analytics Applications*. 2015. Procedia Computer Science.
2. Altman DG, Vergouwe Y, Royston P. (2019). *Prognosis and prognostic research: validating a prognostic model*. BMJ 2019; 338: b605.
3. Aly Megahed, Guang-Jie Ren, Michael Firth. (2015). *Modeling Business Insights into Predictive Analytics for the Outcome of IT Service Contracts*. 2015. IEEE International Conference on Services Computing.
4. Ayyanathan N., A. Kannammal. (2015). *Combined forecasting and cognitive Decision Support System for Indian green coffee supply chain predictive analytics*. 2015. IEEE.
5. Chandra J, DrNachamai.M, DrAnitha S Pillai. (2015). *An Integrated Approach to Predictive Analytics for Stock Indexes and Commodity Trading Using Computational Intelligence*. Vol. 5, No. 8, 142-148, 2015. World of Computer Science and Information Technology Journal (WCSIT).
6. Ebenso B, Otu A. (2020). *Can Nigeria contain the COVID-19 outbreak using lessons from recent epidemics?* The Lancet Global Health. 2020.
7. Foteini G., Isidoros P., Ioannis H. (2014). *Utilizing Semantic Web Technologies and Data Mining Techniques to Analyze Students Learning and Predict Final Performance*. December 2014. International Conference of Teaching, Assessment and Learning.
8. Geisser D., Seymour H. (2003). *Predictive Inference: An Introduction*. Chapman & Hall. p. [page needed]. ISBN 978-0-412-03471-8.
9. Hina Gulati. (2015). *Predictive Analytics Using Data Mining Technique*". 2015. IEEE
10. Hoda Moghimi, Stephen Vaughan, Steven McConche, Nilmini Wickramasinghe. (2016). *How Do Business Analytics and Business Intelligence Contribute to Improving Care Efficiency?*. 2016. 49th Hawaii International Conference on System Sciences.
11. Ibrahim JG, Chu H, Chen MH. (2016). *Missing data in clinical studies: issues and methods*. J Clin Oncol 2016; 30: 3297–3303.
12. Liaw A, Wiener M. (2014). *Classification and regression by random Forest*. R News 2014; 2: 18–22.
13. MacLean, O.A., Orton, R., Singer, J. B., Robertson, D. L. (2020). *On the origin and continuing evolution of SARS-CoV-2*. http://virological.org/t/response-to-on-the-origin–and-continuing-evolution-of-sars-cov-2/418.
14. Meryem Ouahilal, Mohammed El Mohajir, Mohamed chahhou, BadrEddine El Mohajir. (2016) *A Comparative Study of Predictive Algorithms for Business Analytics and Decision Support systems: Finance as a Case Study*. 2016. IEEE
15. Sahin, A. R. (2020). *Novel Coronavirus (COVID-19) Outbreak: A Review of the Current Literature*. Eurasian Journal of Medical Investigation, 4(1), 1–7. https://doi.org/10.14744/ejmo.2020.12220
16. Siegel CA, Siegel LS, Hyams JS. (2021). *Real-time tool to display the*

predicted disease course and treatment response for children with Crohn's disease. Inflamm Bowel Dis 2021; 17: 30–38.

17. Singal AG, Rahimi RS, Clark C et al. An automated model using electronic medical record data to identify patients with cirrhosis at high risk for readmission. Clinical Gastroenterol and Hepatol 2013; 11: 1335–1341.
18. Steyerberg EW, Harrell FE Jr, Borsboom GJ. (2021) *Internal validation of predictive models: efficiency of some procedures for logistic regression analysis*. J Clin Epidemiol 2021; 54: 774–781.
19. Steyerberg EW, Vickers AJ, Cook NR. (2020). *Assessing the performance of prediction models: a framework for traditional and novel measures.* Epidemiology 2020; 21: 128–138.
20. Seung-Jun Shin, Jungyub Woo, SudarsanRachuri. (2014). *Predictive analytics model for power consumption in manufacturing*. 2014. Elsevier B.V.
21. Toll DB, Janssen KJ, Vergouwe Y. (2018). *Validation, updating and impact of clinical prediction rules: a review*. J Clin Epidemiol 2018; 61: 1085–1094
22. Usamah bin Mat, Norlida B., Pauziah M., Rosni Abu K.. (2013). *An Overview of Using Academic Analytics to Predict and Improve Students' Achievement: A Proposed Proactive Intelligent Intervention*. 2013. IEEE 5th Conference on Engineering Education.
23. Uzma Afzal, Tariq Mahmood. (2013). *Using Predictive Analytics to Forecast Drone Attacks in Pakistan*. 2013. IEEE
24. Waljee AK, Joyce JC, Wang SJ. (2010). *Algorithms outperform metabolite tests in predicting response of patients with inflammatory bowel disease to thiopurines*. Clin Gastroenterol Hepatol 2010; 8: 143–150.
25. Waljee AK, Higgins PD. (2020). *Machine learning in medicine: a primer for physicians*. Am J Gastroenterol 2020; 105: 1224–1226.
26. Wiki(2020):http://en.wikipedia.org /wiki/2020_coronavirus_pandemic s_in_Nigeria. Accessed April, 19 2020.
27. Woo, P. C. Y., Huang, Y., Lau, S. K. P., & Yuen, K. Y. (2010). *Coronavirus genomics and bioinformatics analysis*. Viruses, Vol. 2, pp. 1805–1820.https://doi.org/10.3390/v20 81803
28. Yafeng Lu, Robert K., Dennis Thom, Feng Wang, Steffen Koch, Thomas Ertl, Ross Maciejewski. (2014). *Integrating Predictive Analytics and Social Media*. November 9- 2014. IEEE.