International Journal of Basic and Applied Science 10 (3) (2022) 96-102

Published by: IOCSCIENCE

International Journal of Basic and Applied Science
Journal homepage: www.ijobas.pelnus.ac.id

A Look at the extrapolated losses due to fire outbreak in Anambra State, Nigeria

Aronu, C. O.[1*], Bilesanmi A. O.[2] and Arowolo, O. T[3] and Agbogidi, R. B.[4]
[1]Department of Statistics, Chukwuemeka Odumegwu Ojukwu University, Uli, Anambra State, Nigeria
[2]Department of General Studies, Petroleum Training Institute, Effurun-Delta State, Nigeria
[3]Department of Mathematical Sciences, Lagos State University of Science and Technology, Ikorodu-Lagos, Nigeria
[4]Mathematics Units, School of Basic Science, Nigeria Maritime University, Okerenkoko, Delta State

Article Info

Article history:

Received Aug 25, 2022
Revised Nov 20, 2022
Accepted Dec 11, 2022

Keywords:

Economy
Fire Disaster
Property
Regression
Random Forest

ABSTRACT

The recent rise in fire incidents in Anambra State has resulted in displacement, stress, psychological, adverse effect on the lives of its inhabitants and a devastating consequence on the economy. The objectives of the study include examining the estimated value of properties loss due to fire outbreak within the study period and to employ the Random Forest regression model to predict the estimated losses associated with fire disasters in Anambra State. The data for the study was a secondary data obtained from the records Department of the Anambra State Fire Service, Headquarters Awka. The study used the Random Forest regression method to analyze the data obtained in this study. The Random Forest regression analysis was employed to predict the estimated value of properties loss (EVPL) due to fire outbreak. The explanatory variables used for the prediction of the response variable were Number of victims (NV), Percentage of Plain Land (PL), Population Size (PS), Population Density (PD) and Actual Revenue by LGA (ARLGA). The findings of the study revealed that EVPL has a higher Skewness and Kurtosis followed by NV and the least was found to be PL. Further result revealed that the percentage of variance explained was 12.03%, sum of square error (SSE) was 31.41, root mean square error (RMSE) was 0.7236 while the R-square was 59.18%. This result implies that the model was moderately positively adequate since it recorded a positive coefficient of determination. The findings from the variable importance analysis showed that Population Density played a major role in the estimation of the response variable followed by the number of victims while population size was found to be the least important variable for estimating the response variable.

Corresponding Author:

Aronu Charles O.,
Department of Statistics, Chukwuemeka Odumegwu Ojukwu University,
Uli, Anambra State, Nigeria.
Email: amaro4baya@yahoo.com

1. INTRODUCTION

The growing trend of fire epidemics in Anambra state has called for urgent attention from the government and the citizens. It is a shared responsibility to protect and ensure the safety of lives of persons residing in the State, hence the need to call attention to the dangers posed by fire by way of data analysis. Fire is a multifaceted concept whose meaning varies depending on the context and interpretations of individual scientists. Fire can be defined as the process of oxidation of combustible materials that results in the generation of heat, lightning, and products such as smoke and carbon dioxide [1]. Not every fire, though, is disastrous for people's lives, property, and infrastructure. A typical fire is actually a good source of heat energy for indoor and outdoor uses. A fire is only considered catastrophic if it substantially outpaces the capacity of the afflicted society to respond to it using its own resources, causing significant human and material losses that undermine society's ability to operate. When fire hazards and vulnerabilities come together, fire disasters happen [1]. Events known as "fire hazards" can have a detrimental impact on a community [2]. It may be formed either naturally or intentionally [1]. While man-made fire hazards result from human actions like electrical short-circuiting, cooking, and arson, natural fire hazards are caused by natural factors like lightning, earthquakes, and extremely high temperatures [1]. The term "fire vulnerability" refers to a person or community's susceptibility to fire hazards [2].

Fires might be characterized as unwelcome fires considering that it is difficult to predict where and when a fire catastrophe will occur. Everyone engaged is compelled to foresee unplanned fire disasters and act as quickly as possible to prevent them. A fire in a building is an unpleasant disaster because it affects people, property, and the environment outside the structure. Fires can potentially result in a high rate of fatalities and material losses, so it is important to pay attention to everyone's safety, including workers within buildings and nearby neighbours. One of the issues in urban settings that are challenging to avoid is fire in structures. Poor urban fire fighting infrastructure and accessibility, the use of wood and charcoal for room heating, poverty, ignorance, and burning yard garbage are common causes of fires in urban areas [3]. Other common causes include flaws in disaster preventive programs. One of the key concerns of urban planners is the frequency of fire disasters, which are frequently occurring in congested central business districts (CBDs) and markets. Congestion and a high population density enhance the likelihood of fire hazards, raising the probability of a fire and the need for rescue operations. The characteristics of areas that are at risk for fire dangers include dense population density, erratic building patterns, and low-quality buildings, as well as a shortage of fire fighting supplies. The linear distance between the houses will make it difficult for the firemen's vehicles, and the failure of the fire hydrants will make it easier for the fire to spread. Fire outbreaks in private and public buildings are a problem in Nigeria and have led to the loss of life and the destruction of property [4]. Considering these fire outbreaks in Anambra State, Nigeria, it is therefore important to conduct a study that can review the estimated loss due to fire outbreaks.

In Anambra State, the Fire Service Departments of both the Federal and State are saddled with the responsibility of managing fire mishaps in the state while protecting lives. However, the frequent occurrences of fire accidents can be linked to displacement, stress, psychological effects, detrimental effects on residents' quality of life in Anambra, and adverse effects on the state's economy. The crucial question is whether the state government's recent overhaul of the Anambra State Fire Service in coordination with the federal government has considerably helped to reduce the emergence of fire disaster patterns in the state. Although it is well recognized that protecting and ensuring the safety of people living in the State is a shared responsibility, there is still a need to draw attention to the risks that fire poses.

2. RESEARCH METHOD

2.1 Study Area

Anambra state is one of the eminent states in the south-eastern part of Nigeria. Its original name 'Oma Mbala' is the native name of the Anambra River. Its capital is in Awka. Nnewi, Onitsha and Ekwulobia are the biggest commercial hubs and industrial cities in the State. The state shares boundaries with Delta State, Imo State, Rivers State, Enugu State and Kogi State. The indigenous ethnic groups in Anambra state are the Igbo (98% of its population) while the Igala's consist of 2% of its population.

Anambra is one of the most populated and densely populated states in Nigeria after Lagos State. It stretches between Amorka and Oba (45 km) with an estimated average density of 1,500–2,000 persons/sq.km. The state has an annual population growth rate of 2.83% per annum, with more than 60% of people living in the urban area of the State.

2.2 Method of data collection

A secondary data was collected from the Anambra State Fire Service, headquarters Awka, dated from

2011 to 2019. This data consists of reported fire incidents from the various fire stations in Anambra State. The variables considered in the study include: Estimated Value of properties loss (EVPL), Number of victims (NV), Percentage of Plain Land (PL), Population Size (PS), Population Density (PD) and Actual Revenue by LGA (ARLGA).

3.3 Method of Data Analysis

3.3.1 Random Forests

An ensemble learning technique is used for classification and regression in the supervised learning algorithm known as random forest [5]. It is not a boosting approach; rather, it is a bagging technique. Random Forests functions by building a large number of decision trees during training time and then determining the class that corresponds to the mean prediction (for regression) or mode of classes (for classification) of the individual trees. The parallel operation of the trees in random forest suggests that there is no interaction between them while they are growing. Overfitting is avoided by Random Forests. It covers a wide range of qualities and aids in locating the crucial characteristics. It has two parameters that are easy to use: *ntree (J)* and *mtry (m)*.

***ntree (J)*:** There are thus many trees in the forest. n tree is set to 500 trees by default. In contrast to other ensemble approaches, generalization error initially decreases as J rises; nevertheless, eventually, J grows to an unmanageable size and over fitting takes hold, leading to an increase in generalization error. With Random Forests, this is different. In his work, [5] demonstrated that the generalization error for Random Forests converges almost certainly to a limit as J rises. This means that J can be selected to be as large as desired without bothering about the generalization error rising. The only significant issue with J is that it shouldn't be too low because low values of J can lead to unstable and erroneous out-of-bag estimates. When J is large enough, the estimated generalization error can be used to determine when it has stabilized.

***mtry (m)*:** This is the total number of predictor variables that were chosen at each node at random. In classification, the default value for *m* is p (rounded up), and in regression, it is [p/3] (rounded down) (s). In his work, [5] advocated attempting the default, half of the default, and twice the default before choosing the best when choosing *mtry*.[6]

Data Splitting: This entails splitting the input data into two data sets—training and test sets—at a given probability distribution. It aids in preventing over fitting and enhancing the precision of the training data set. The training data set and the test data set was created by dividing the research's data into two groups with probabilities of 0.7 and 0.3, respectively.
Training Data Set: This data set of examples is utilized to suit the parameters throughout the learning process. It is the sample of data used to fit the model.
Test Data Set: Although separate from the training data set, this data set has the same probability distribution as the training data set. It is utilized to offer a fair assessment of the final model fit on the training data set.
For any given training dataset

$$D=\left\{\left(x_1, y_1\right), \cdots, \left(x_N, y_N\right)\right\} \text{ with } x_i=\left(x_{i1}, \cdots, x_{ip}\right)^T, \text{ for } j = 1 \text{ to } J,$$

random forest performs regression task and the algorithm for random.
Forest regression looks like this:

1. Take samples from the training dataset for *ntree* bootstraps.
2. For each of the bootstrap samples, construct an unpruned regression tree with the change that, at each node, the best split should be selected from a randomly selected subset of the predictors rather than the best split among all predictors.
3. By combining the predictions of the *ntree* trees, forecast fresh data.

Consequently, the following equation is the model for forecasting a random forest regression utilizing the research data (training data set) at a new point x.
Therefore, the model for predicting a random forest regression using the research data (training data set) at a new point x is given by;

$$\hat{y} = \hat{f}(x) = \frac{1}{j}\sum_{j=1}^{J}\hat{h}_j(x) \quad (1)$$

where $\hat{y}$ is the observations on each of the response variables (i.e. Estimated value of properties loss (EVPL)) in the training data set *D*,
$\hat{f}$(x) is the predicted values of predictor variables (Percentage of Plain Land (PL), Population Size (PS), Population Density (PD), and Actual Revenue by Local Government) of the training data sets used for the research,
J = 500 is the number of trees in the forest, and $\hat{h}_J$(x) is the prediction of the response variable at x using the jth tree, of the training set D.

The mean of squared residuals is computed as

$$MSE_{OOB} = \frac{\sum_{i=1}^{n_{out-of-bag}} \left(y_i - \hat{y}_{i,\text{out-of-bag}}\right)^2}{n_{out-of-bag}} \quad (2)$$

Where, y_i is the value of the response variable for row i ,
$\hat{y}_{i,\text{out-of-bag}}$ is the average of the out - of - bag predictions for the ith observation,
$n_{out-of-bag}$ is the number rows that appear in the out-of-bag data over the entire forest
Reduced mean square error (MSE) and node purity are the two outcomes of the random forest regression. The foundation of MSE is the permutation of data from a bag for each tree and predictor, followed by the averaging of errors. Node purity is the overall averaged reduction in the residual sum of squares when splitting on a variable (i.e. how well a predictor decreases variance). MSE is a more trustworthy indicator of variable importance than Node purity, thus if the two-important metrics (MSE and Node purity) produce conflicting findings, pay attention to MSE.
On the other hand, coefficient of determination, R^2 is the proportion of the total variance in y that has been explained by the regression model is given by; MSE_{OOB}

$$R^2 = 1 - \frac{MSE_{OOB}}{\sigma_y^2} \quad (3)$$

where σ_y^2 is computed with n as divisor (rather than n - 1).
It should be noted that equations without a constant term can have an R-square value that is negative. Since the definition of R-square is the percentage of variance explained by the fit, it is possible for the fit to be worse than just fitting a horizontal line, in which case R-square will be negative. R-square cannot be understood as the square of a correlation in this situation. Such circumstances suggest that a constant term has to be included in the model.

The "**percentage variance explained**" is computed as
% var. explained = (R^2) × (100

Root Mean Square Error (RMSE): This is the mean squared error's square root. One method for determining how well a regression model fits a dataset is to use this method. A given model can "fit" a dataset more accurately the lower the RMSE. The root mean square error, or RMSE, is calculated using the following formula:

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (4)$$

Variance Importance: This measure of predictor variable relevance is helpful in choosing variables and understanding the fitted forest. Permutation importance is used by random forests to do this task. The permutation significance is a metric for assessing the accuracy of predictions when variables are randomly permuted from samples taken from a bag. Given that it is independent of internal model parameters like linear regression coefficients, this approach has a wide range of applications (which are really poor proxies for feature importance). According to intuition, the permutation-based relevance of variable K is a calculation of the prediction error or MSE on a test set that would result from randomly permuting the value of the variable K inside the test set [6].

3. RESULTS AND DISCUSSIONS

This section presents the results of the analysis using the data obtained for the study which available on request.

3.1 Descriptive Summary of the Variables

This section presents the descriptive analysis of the data obtained for the study

Table 1. Descriptive Summary result of the variables

	EVPL	NV	PD	PL	PS	ARLGA
Mean	97139648	4094.238	2284.952	36.76190	267958.4	1.67E+08
Median	1800000.	0.000000	1485.000	40.00000	229400.0	28018270
Skewness	7.290437	5.525842	2.079588	-0.463392	1.330090	2.677263
Kurtosis	59.28935	34.12597	6.408654	2.637276	4.220608	12.25481
Number of Observations	105	105	105	105	105	105

The result obtained in Table 1 shows that the mean value of EVPL was N97139648, NV was 4094.238, PD was 2284.952, PL was 36.76%, PS was 267958.4 and ARLGA was N167, 000,000. The Skewness was obtained for EVPL as N7.29, NV as 5.52, PD as 2.07, PL as -0.46%, PS as 1.33 and ARLGA as N2.68. Also, the Kurtosis was obtained EVPL as N59.29, NV as 34.13, PD as 6.41, PL as 2.64%, PS as 4.22 and ARLGA N12.25. This result revealed that EVPL has the higher Skewness followed by NV and the least was found to be PL while it was revealed that EVPL has a higher Kurtosis followed by NV and the least was found to be PL.

3.2 Random Forest Regression Analysis

This section presents the result of the random forest regression analysis for predicting the estimated value of properties loss (EVPL)

Table 2. Summary result of the Random Forest Regression Analysis

Number of trees	No. of variables tried at each split	Mean of squared residuals	% Var explained	SSE	RMSE	R-Square
500	1	6.350161e+16	12.03	31.4123031	0.7235595	59.182296

The result presented in Table 2 showed that the percentage of variance explained was 12.03%, SSE was 31.41, RMSE was 0.7236 while the R-square was 59.18%. This result implies that the model was fairly positively adequate since it recorded a positive coefficient of determination.

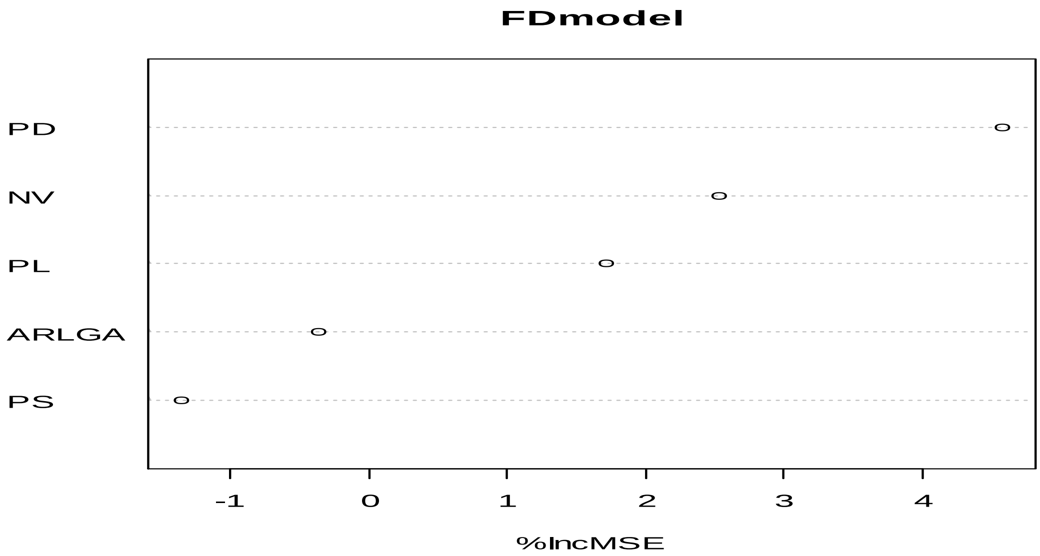

Figure1. Percentage Inclusion of MSE for the predictor variables in the Random Forest Regression Model

The result presented in Fig. 1 shows the variable importance of the predictors in estimating the response variable (EVPL). The result revealed that PD was the most important predictor variables flowed by NV, PL, ARLGA, while PS was found to be the least important variable for estimating the response variable.

4. CONCLUSION

Using a machine learning technique, this study looked at the estimated damages brought on by the fire outbreak in Anambra State. Over the past ten years, Anambra state has experienced repeated fire outbreaks. The state's government has made measures to prevent fires from spreading so that people's homes, businesses, and lives are all protected. The estimated value of property losses (EVPL) as a result of fire breakout was predicted using the Random Forest regression methodology. In order to predict the response variable, the explanatory variables considered were Number of victims (NV), Percentage of Plain Land (PL), Population Size (PS), Population Density (PD) and Actual Revenue by LGA (ARLGA). The findings of the study revealed that EVPL has a higher Skewness and Kurtosis followed by NV and the least was found to be PL. Further result revealed that the percentage of variance explained was 12.03%, SSE was 31.41, RMSE was 0.7236 while the R-square was 59.18%. This result implies that the model was fairly positively adequate since it recorded a positive coefficient of determination. The findings from the variable importance analysis showed that Population Density played a major role in the estimation of the response variable followed by the number of victims while population size was found to be the least important variable for estimating the response variable. The government should make sure that firemen are safe and protected from criminals and vandals. In order to reduce fire accidents in Anambra State, there needs to be better cooperation between the Federal and State Fire Services.

REFERENCES

[1] Sena, L., and Michael, K.W."Disaster prevention and preparedness", Ethiopia: Ethiopia Public Health Initiative, 2006https://pdf.usaid.gov/pdf_docs/pnaec413.pdf.

[2] Nyagawa, Z. M. "An investigation of the immediate causes of fire disasters in boarding secondary schoolshttp://oapub.org/edu/index.php/ejes/article/view/1344

[3] Setiawan, C., Handawati, R., and Ermalia "Analysis of Influence Settlement Densityon the Fire Hazards settlement at Cengkareng Subdistrict", West Jakarta, IOP Conf. Series: Earth and Environmental Science, 2020, 412(012011): 1-9. doi:10.1088/1755-1315/412/1/012011https://iopscience.iop.org/article/10.1088/1755-1315/412/1/012011/meta

[4] Odaudu, U. S. "Analysis of the Overview of the Causes of Fire Outbreaks in Nigerian Markets", CARD International Journal of Environmental Studies and Safety Research (IJESSR), 2017, 2(2): 1-

9.https://www.afrigistjournals.com/index.php/ajlaem/article/view/v1i1a1
[5] Breiman, L. “Manual on setting up, using and understanding random forests”, V3.1 2002. https://oz.berkeley.edu/users/breiman/Using_random_forests_V3.1https://proceedings.neurips.cc/paper/2013/hash/e3796ae838835da0b6f6ea37bcf8bcb7-Abstract.html
[6] Cutler, A., Cutler, D.R. and Stevens, J.R. “Random forests”, Chapter in Machine Learning, 2011. DOI:10.1007/978– 1 – 4419 – 9326 – 7 _ 5. Source: DBLP"https://link.springer.com/protocol/10.1007/978-1-60761-744-0_17