TEST Engineering & Management

November-December 2019
ISSN: 0193-4120 Page No.3913- 3918

A Conceptual Framework for Predicting the Effects of Encroachment on Magnitude of Flood in Foma-River Area, Kwara State, Nigeria Using Data Mining

A.A. Chindo[1], I.N.M. Shaharanee[2], J.M. Jamil[3]

[1,3]School of Quantitative Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia
[2]Othman Yeop Abdullah Graduate School of Business, Universiti Utara Malaysia,
06010 Sintok, Kedah, Malaysia
[1]chindo_abdulrasaq@ahsgs.uum.edu.my, [*2]nizal@uum.edu.my, [3]jastini@uum.edu.my

Article Info
Volume 81
Page Number: 3913 - 3918
Publication Issue:
November-December 2019

Abstract

Flooding occurs but there is no flood hazard. It is only after human encroachment into the floodplain that turn into hazard. The practice of continuous increase of properties development along the floodplain and indiscriminate refuse disposal into water channels have been a major constituting factors to intensive flooding along the floodplain. This is as a result of decline in the capacity of floodplain to absorbs excess flooding, thus resulting to exposing more urban areas to be vulnerable to flood. Foma-river is located in Ilorin Kwara, Nigeria on latitude N08,49574 and longitude E004,5107. Climate of Ilorin comprises of the dry and wet seasons with the wet season starting around March and lasting for about four to five months. This study intends to propose a conceptual framework to support the prediction of effects of season on magnitude of flood in Foma-river area using data mining approach based on 7 years sampled data from Nigeria Meteorological Agency (NIMET), and questionnaire responses from residents along Foma-river floodplain

Article History
Article Received: *5 March 2019*
Revised: *18 May 2019*
Accepted: *24 September 2019*
Publication: *19 December 2019*

Keywords: *flood plain, encroachment, data mining, meteorological, morphological.*

1. Introduction

The likely consequences of climate change are profound, particularly on people in the less advanced countries. Nigeria experiences heavy rainy season between the periods of May and September yearly, this led to a serious and protracted flash floods across the country. Due to poor drainage systems or outright lack of it, some rural areas and overcrowded slums becomes so deadly. Precisely, many coaster and inland areas experienced higher precipitation on the second of July. Coaster city like Lagos were severely flooded, while the flood also got rampaged in some areas like Challenge, Oke-Ayo, and Eleyelein Ibadan. At least 39 people lost their lives to flood towards the tail end of July same year in plateau state, central Nigeria. About 200 houses were underwater When the Lamingo dam overflow and swept across many locations in Jos and at least 35 people were reported missing.

Several cases of flooding of urban cities of Nigeria has been reported in recent times mostly Sokoto, Lagos, Ibadan, Abeokuta, Gusau, and Makurdi, to mention but a few(Alayande, W. A., Mohammed, G., Caleb, I., &Deimodei, M. I.,2012). Excess releases from dams whose functioning capacities could not cope with extreme inflows into their reservoir areas has also been identified as responsible for flooding in Nigeria mostly in the Cities downstream major dams (Alayande et. al, 2012). Foma-river is positioned approximately seven kilometers from the Emir's palace, Ilorin, Nigeria on latitude N08,49574 and longitude E004,5107. The river is quite freely

November-December 2019
ISSN: 0193-4120 Page No.3913- 3918

flowing during the occurrence of rain and partially slow during the dry season (Agbabiaka, T. O., &Oyeyiola, G. P,.2012).

Feasibility studied reflects that the river has its origin at 'WAHRI' along Gerewu, Ilorin west local government of Kwara state. It joined Asa-river at a point called Ibu-Afonja, beneath the Sobi-rock, in Ilorin south area of Kwara state. The catchment area of the river is estimated be around 303,309,421.02 square meters. The river, especially during the dry season overflows its bank which over the years had resulted into several kilometers stretch of floodplains. The human encroachment into the floodplain had triggered several damages to buildings, institutions, and other social facilities in the affected communities. Most of the literature reviewed dealt with the effects of floodplain on flood hydrographs and did not specifically treat the encroachment effects on seasonal magnitude of flood. Some other researches involved the evaluation of some typical encroachment problems in which analysis were made using computer programs such as HEC-1 (Hydrological Engineering Centre 1973), and HEC-2 (Hydrological Engineering Centre 1979), fundamentally, the goals of performing this encroachment analysis is to determine the limits of encroachment that will cause a precise change in surface water elevation. To determine changes in surface water elevation, the program must first determine a natural profile of the river with no encroachment. Otherwise known as 'BASE PROFILE'.

The application of data mining operations in this research was due to the fact that the approach has a relative fewer assumptions compared to statistical approach in managing effects of flood. In statistical approach, concerns such as normality of distribution, collinearity of factor or independence of error terms are seriously a subject of determination. While the data mining techniques have fewer and flexible assumptions which is capable of producing better and accurate estimate, especially when dealing with large data sets (Thammasiri, D., Delen, D., Meesad, P., &Kasap, N.,2014). Data mining operations are unique tools that have been proven to be efficient in handling the missing data, productive in capturing the complexity of nonlinearity, and quite reliable in producing accurate prediction. Despite overwhelming decisions on the use of data mining techniques, the superiority of a particular technique over the others to make prediction cannot be established. The superiority of a technique over the others can better suggested through the nature of the data and composition of the problems in research. The lack superiority of one technique to the other, prompted an investigative procedure to ensure a most desirable data mining technique is apply to a specific prediction problem (Thammasiri.et al, 2014). Therefore, in this research study, we intend to develop a multinomial logistic predictive model using data from Nigerian meteorological agency and responses from households' questionnaire.

We shall also construct a 6 X 6 confusion matrix, based on the data presentation for magnitude of flood which are classified as; 1= extreme flood; 2 = severe flood; 3 = moderate flood; 4 = mild flood; 5 = normal flood; and 6 = no flood. In other to derive the measurement metrics for models' comparison, we make use of 10-fold cross validation. Our experiment will be based on 2 x 10 design that will produce 20 classification models.

2. Related Works

There is no question about whether climate change is happening, but what to do about it. Over the last 20 years climate change has become an ever more high-profile issue both from social and economic viewpoint. Understanding that climate is slowly and steadily changing due to human activities, and natural disasters. Flooding affects more people on an annual basis than any other forms of natural disasters (Kolawole, O. M., Olayemi, A. B., & Ajayi, K. T. ,2011). Ilorin is the capital of Kwara State, Nigeria located in western Africa. Ilorin city lies between latitude $8^0$24'N and $8^0$36'N and between longitude $4^0$10'W and $4^0$36'E. The city is drained mostly by consequent River Asa, and its tributary streams such as Aluko, Alalubosa, Okun, Osere, Agba&Alileke (Agbabiaka&Oyeyepo. 2012).The period of wet season in Kwara state extends between March and October. On the other hand, the dry season falls around November up to March ending.

Floodplain act as a barrier, garden and an important area for organic life. The area provides fresh water for wet land and an important environment around the river or stream (Ndabula G, G. Jidauna, K. Oyatayo, P.D. Averik, & E.O. Iguisi,.2012). The floodplain areas are the most environment vulnerable to flooding, especially during excess rain, reservoirs collapse or surface run-off. The floodplains of major rivers are giving ways to early urbanization (Ndabula et al, 2012). The compositional characteristics of floodplains along the major rivers which indeed blessed with fertile soil, flat landscape and moldy surface make them less expensive to acquire. The floodplain areas are less expensive, thus attract very many low-income earners, farmers, fishermen and led to encroachment which has increases risk of flood and environmental hazards. The emergent of high socio-economic interest in this environment is adjusting the incorporation of surface-runoff with the main networks, reducing the merging capability and surface water storage. This actions also leading to downstream pollution, decreases the quality of water and actual water resources (Ramachandra, T. V., Aithal, B. H., & Kumar, U.,2012). By extension, the concept of urban encroachment in floodplains has heightened the extension of internal flooding which can attributed to reduction in the capacity of floodplain to check flood, thus exposing lots of urban residential vulnerable to flood.

November-December 2019
ISSN: 0193-4120 Page No.3913- 3918

The problems of encroachment into floodplain is making the management of urban floodplain a serious concern all over the World, specifically the modern-day rising trends in urban flooding. This problem is more concern to governments across developing nations where the resources and capacity to implement the floodplain ordinance is lacking (Ndabula et al, 2012). Many residential buildings and shops were underwater in some parts of Ilorin, Kwara State following a heavy rainstorm. The heavy rain lasted about three hours and was accompanied by flooding. The asphalt on some township roads were also washed away by the flood (Nigerian Tribune, 2017).

3. Methodology

In this study, we propose to utilize multinomial logistic regression—along with two data sources— responses from questionnaire, and flood data from NIMET---to build prediction models. Also, a 6 X 6 confusion matrix will be constructed to derive the measurement metrics. The 10-fold validation will be employed to cross examine the measurement metrics. The data set is partitioned into 10-fold, such that the classification exercise is carry out along the partitioned data. The process of investigation will involve the use of nine subset of the cross-classified data as training, while the one part be using for testing. These processes will be repeated ten times for each of the two model. The results of the test will reflect the unbiased performance of the model.

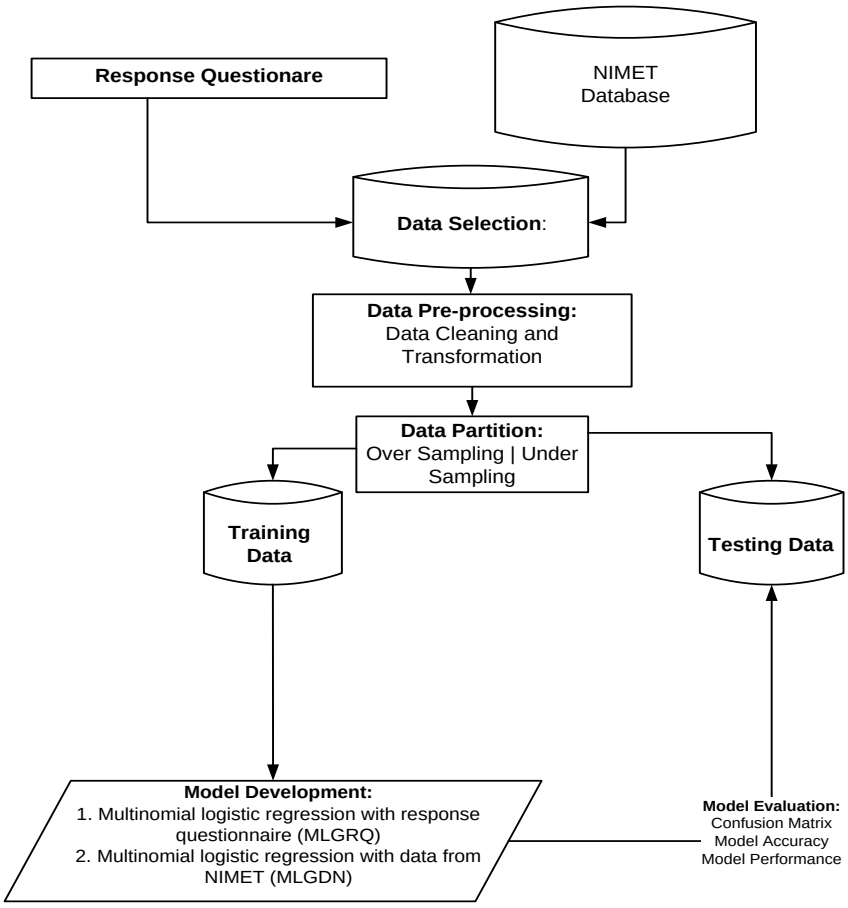

Figure1: The overview of the methodology employed in this study

Data Mining Techniques

The emergence of Information Technology has resulted into numerous availabilities of various forms of databases. While the approach to further store and operate this valuable data to provide practicable decision is now being focused (Fayyad, Piatetsky-Shapiro et al. 1996). The process of extracting useful information from data otherwise known as; knowledge discovery, pattern analysis, mining knowledge from data, are all procedures in data mining. Data mining explored its processes to search through a huge data to find useful information which are perceived to be previously unknown. These discovered information patterns can now be further employed for decision making. Figure 1 exhibits the proposed framework for predicting the effects of encroachment on magnitude of flood in Foma river, Nigeria.

November-December 2019
ISSN: 0193-4120 Page No.3913- 3918

Data Extraction and Collection

The Nigerian meteorological agency (NIMET), is a government body set up to provide monthly information on rainfall, flood, drought and other related assignments. The agency was established through the Act of the National Assembly – "NIMET Establishment ACT" of 2003, which was endorsed on 21st of the same month and was sign into law by the Nigeria president on 19th June 2003. The responsibilities of this agency among others are; 1) To make available and interpret policies of government in relation to meteorology; 2) To provide climate forecast which can be used by aircraft operations and ocean vessels.; 3) To observe, collate, process, and publicized meteorological data both within and outside Nigeria; 4) To carry out research and paper publication in areas of meteorology to support the sustainable socio-economic development activities in Nigeria. The research is focus collecting spatial flood data between 2012 and 2018. A questionnaire is a research device consisting of a series of questions and other prompts for assembling information from respondents. Although they are often intended for statistical analysis of the responses, this is not always the case. It is use for research work by scientists, businesses, and political parties among others. (Casiey, D. J., &Lury, D. A. (1981).

Table1: Summary of Data Sources

Source	Period	Attributes Name	Data Type	Station Information
Questionare	5 Years	Flood Livelihood Housing Health Water and Sanitation Properties Agriculture Education Magnitute of Flood Other	Nominal	Respondent: One hundred and twenty household in and around the river floodplain
NIMET	5 Years	Flood Report	Nominal	IIorin Metropolis

Data Cleaning and Transformation

The collected data will be analyzed and processed to get the most suitable and relevant data. The raw data in heterogonous formats (word, pdf, excel, web) and may need to extract into manageable records. The data will be evaluated through data completeness, whereas we will use other application such as Gretl 3.0 to obtain regression coefficients and predicted value of confusion matrix. So, as the expected values are predictable. Then, the data from different sources will integrate in a logical method to eliminate data redundancy and distinguish any value conflicts Aziz.et al (2016).

Model Development

We intended to do models comparison for the two sources of data using the following description; multinomial logistic regression with response questionnaire (MLGRQ) and multinomial logistic regression with data from NIMET (MLGDN). The logistic Regression application inspects the non-linear association between the categorized or continuous predictor variables and binary outcome. The application predicts a probability of outcome between 0 and 1 in odds ratio. In essence, it models the probability of an outcome based on their characteristics, simply because the chance is ratio. Thus, the actual chance in the algorithm is define as:

$$\text{Log}\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots\ldots + \beta_m X_m \quad \ldots\ldots\ldots\ldots(1)$$

The regression equations will be more than one when the dependent variable values are more than two. Generally, the regression equation will be one less than the number of outcomes (Morgan and Teachman 1988). In this case, the interpretation can be more complexed because each independent variable is quite related to several regression coefficients. The better procedure to get rid of the challenge is to take cognizance of what each regression is predicting. The basic understanding of this step will enable to give prescription to regression for each variable (Lyness & Heilman 2006).

Model Comparison and Evaluation

To evaluate the performance of the two the models, we used several general metrics. The confusion matrix has proposed in Table 2, will be used to obtain the measurement metrics. The confusion matrix is expected to come out with correct and incorrect outcome for each class. The binary classification problems have four cells; 1) True Positives (TP)— is the number of positive examples that were correctly predicted; 2) True Negatives (TN)— is the number of negative examples that were correctly predicted; 3) False Positives (FP)— is the number of positive examples that were incorrectly predicted; while 4) False Negatives (FN)— is the number

TEST Engineering & Management

November-December 2019
ISSN: 0193-4120 Page No.3913- 3918

of negatives examples that were incorrectly predicted (see Table 2).

To compute predicted value ($\hat{P}$) of the binary dependent variable Y = 0,1. The value $\hat{P}$ must lies between 0 and 1. If $\hat{P}$ = P (Y = 1) , So that;

$0 \leq \hat{P} \leq 1$ ……..(1)

Thus setting the threshold for $\hat{P}$, call it $\bar{P}$. Such that:

$\hat{Y}$ = 1, if and only if $\hat{P} \geq \bar{P}$. Where $\bar{P}$ = 0.5

$\hat{Y}$ = 1, if and only if $\hat{P} \geq$ 0.5 ……(2)

$\hat{Y}$ = 0, if and only if $\hat{P}$< 0.5 ……(3)

Table 2: A typical confusion matrix for a binary classification problem

OBSERVED Y	**PREDICTED $\hat{Y}$**	
	$\hat{Y}$ = 0, Iff $\hat{P}<\bar{P}$	$\hat{Y}$ = 1, Iff $\hat{P} \geq \bar{P}$
0	Y = $\hat{Y}$	Y ≠ $\hat{Y}$
1	Y ≠ $\hat{Y}$	Y = $\hat{Y}$

Similarly, we can constuct multiple confusion matrix for dependent variable Y after runing a multinomial logistic regression in gretl. Gretl 3.0, will automatically set the threshold value for $\hat{P}$, $\bar{P}$ = 0.5. Thus based on the threshold probabity, the predicted values for Y (**$\hat{Y}$),** will be generated.

Table 3: A multiple 6 x6 dimension confusion matrix

	PREDICTED $\hat{Y}$					
	$\hat{Y}$=1,$\hat{P} \geq \bar{P}$	$\hat{Y}$=2,$\hat{P} \geq \bar{P}$	$\hat{Y}$=3,$\hat{P} \geq \bar{P}$	$\hat{Y}$=4,$\hat{P} \geq \bar{P}$	$\hat{Y}$=5,$\hat{P} \geq \bar{P}$	$\hat{Y}$=6,$\hat{P} \geq \bar{P}$
	1	2	3	4	5	6
1	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}
2	Y_{21}	Y_{22}	Y_{23}	Y_{24}	Y_{25}	Y_{26}
3	Y_{31}	Y_{32}	Y_{33}	Y_{34}	Y_{35}	Y_{36}
4	Y_{41}	Y_{42}	Y_{43}	Y_{44}	Y_{45}	Y_{46}
5	Y_{51}	Y_{52}	Y_{53}	Y_{54}	Y_{55}	Y_{56}
6	Y_{61}	Y_{62}	Y_{63}	Y_{64}	Y_{65}	Y_{66}

But in our case, the multiple class classification nature of the dependent variable requires a structured 6 X 6 confusion matrix to be used in performance measure and model's comparison. This research work will be evaluating how good the classifier is at predicting the class label using the following measurement metrics:

Accuracy: Is refer to recognition rate of the classifier, given by percentage of observation that are correctly classified by the classifier (Freitas, 2013). Is derive by dividing the total number of correct predictions by the total number of the observation.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} = \frac{TP+TN}{N}$$

F1-Score: Is harmonic mean of the recall and precision and purnishes extreme values more. So, it obtained the harmonic average thus stimulate disparity in an imbalanced data.

$$\text{F1-Score} = \frac{2^{*}Precision\ ^{*}Recall}{Precision\ +Recall}$$

Precision: Is a measure of exactness of the classifier. Given a class prediction from the classifier, how likely is it to be correct?

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall: Is a measure of completeness of a classifier. It reflects the proportion of positive class that are correctly classified. Given a class, will the classifier ditect it? The same as sensitivity.

$$\text{Recall} = \frac{TP}{TP+FN}$$

Acknowledgments

This work was supported by MOE under Fundamental Research Grant Schema.

References

[1] Agbabiaka, T. O., &Oyeyiola, G. P. (2012). Microbial and physicochemical assessment of Foma River, Ita-Nmo, Ilorin, Nigeria: an important source of domestic water in Ilorin Metropolis. *International Journal of Plant, Animal and Environmental Sciences*, *2*(1), 209-16.

[2] Alayande, W. A., Mohammed, G., Caleb, I., &Deimodei, M. I. (2012). Assessment of urban flood disaster: a case study of 2011 Ibadan floods. *Special Publication of the Nigerian Association of Hydrological Sciences*, 13-23.

[3] Aziz, A. A., Harun, N. A., Makhtar, M., Abdullah, F. S., Jusoh, J. A., & Zakaria, Z. A. (2016). A conceptual framework for predicting flood area in terengganu during monsoon season using association rules. *Journal of Theoretical & Applied Information Technology*, *87*(3).

[4] Casiey, D. J., &Lury, D. A. (1981). *Data collection in developing countries* (No. 519.59 C3).
[5] Drought and flood monitor bulletin. Nigerian meteorological Agency, 2013-2017
[6] Flood damages properties in Ilorin. Nigerian Tribune: 28th May 2017.
[7] Hydrological engineering centre (1973), 'HEC-1 flood Hydrograph package,' user's manual computer program 723-X6-L2010. U.S Army corps of engineers.
[8] Hydrological engineering centre (1979), 'HEC-2 water surface profil,' user's manual computer program 723-X6-L202A. U.S Army corps of engineers.
[9] Kolawole, O. M., Olayemi, A. B., & Ajayi, K. T. (2011). Managing flood in Nigerian cities: Risk analysis and adaptation options–Ilorin city as a case study. *Archives of Applied Science Research*, *3*(1), 17-24.
[10] Ndabula G, G. Jidauna, K. Oyatayo, P.D. Averik, and E.O. Iguisi: Analysis of Urban floodplain encroachment: Strategic Approach to flood and floodplain Management in Kaduna Metropolis.
[11] Ramachandra, T. V., Aithal, B. H., & Kumar, U. (2012). Conservation of wetlands to mitigate urban floods. *Journal of Resources, Energy and Development*, *9*(1), 1-22.
[12] Thammasiri, D., Delen, D., Meesad, P., &Kasap, N. (2014). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, *41*(2), 321-330.
[13] Fayyad, U., et al. (1996). "From data mining to knowledge discovery in databases." AI magazine 17(3): 37.
[14] Lyness, K. S. and M. E. Heilman (2006). "When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers." Journal of Applied Psychology 91(4): 777.
[15] Morgan, S. P. and J. D. Teachman (1988). "Logistic regression: Description, examples, and comparisons." Journal of Marriage and Family 50(4): 929-936.