ZAMBIA INFORMATION COMMUNICATION TECHNOLOGY (ICT) JOURNAL

Volume 6 (Issue 1) (2022) Pages 18-24

Predicting Climate Change Related Extreme Natural Disasters Using Machine Learning in Zambia.

David Phiri[a] and Christopher Chembe[b]

a. Department of Information and Communication Technology Studies, National Institute of Public Administration (NIPA), Lusaka, Zambia, E-mails: phirid38@gmail.com

b. School of Computing, Technology and Applied Sciences, ZCAS University, Lusaka Zambia E-mail: christopher.chembe@zcasu.edu.zm

***Abstract* — *One* of the most important concerns affecting humanity today is climate change that has led to increased frequency of natural disasters that threaten social and economic stability to populations. Zambia's vulnerability to the threat of disasters remains high because the country still lacks an effective Early Warning System (EWS). This study recognises the need to evaluate various Machine Learning (ML) algorithms, that have been successfully implemented in disaster prediction, in order to develop a model for Zambia. Six ML algorithms, namely; Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbor (KNN), Gaussian Naive Bayes (GNB), Decision Tree (DT), and Support Vector Machine (SVM), have been compared from which the best performing is chosen. The historical climate data is obtained from the Zambia Meteorological Department (ZMD) while historical natural disasters data was obtained online because it is not locally available. The study results show that LR and SVM algorithms performed better than the others, both scoring 73.0% accuracy, respectively. LR is chosen to produce the final model because it has a shorter computational time compared to SVM. The model is then incorporated in a web service and android application for deployment. However, the high number of outliers, missing values and highly imbalanced classes affect the performance of the model. ML data cleaning and feature engineering techniques, such as *Data Imputation* and *Oversampling Techniques*, are applied but certain challenges still persist because these tools have their own flaws. Therefore, the model's performance in a real-world data environment is likely to be affected.**

***Keywords* — Climate Change, Early Warning Systems, Machine Learning, Natural Disasters, Prediction Model**

1.0 Introduction

Climate change is one of the most significant issues facing humanity today. These issues include the escalating frequency of natural disasters brought on by climate change, such as droughts, floods, forest fires, heat waves, tropical cyclones, storms, tsunamis, avalanches, tornadoes, severe thunderstorms, and hurricanes [1]. As a direct consequence, climate change catastrophes threaten infrastructure, social and economic circumstances, human and animal health, energy, agriculture, and the natural environment, among other things [2].

Being part of the international community, Zambia is not immune to the consequences of climate change. According to studies, during the past forty years, the frequency of catastrophic occurrences like floods and droughts have increased [3]. Given this situation, the nation needs an Early Warning System (EWS) that is effective and efficient and can help forecast major disaster occurrences brought on by climatic variability.

Zambia largely relies on weather forecasting that in turn is used to predict the occurrence of disasters. However, this approach seems not to be effective hence the country has continued to experience high losses in natural disasters due to the failure of this approach to accurately and timely forecast disaster incidences. Aside from a possible lack of competent human resources to run EWSs and the cost associated with their implementation, the assumption in this study is that the other cause could also be the failure of technological tools, such as climate disaster prediction models, being employed.

Machine Learning (ML), a subfield of computer science, has grown in popularity recently as it provides flexible tools for technological advancement [4]. Its ideas have been promoted as helpful instruments for mitigating climate change effects, particularly in creating forecasts for climate change-related occurrences like catastrophes.

The main purpose of this research project is to identify gaps in the current EWSs in Zambia and propose the application of ML approaches that can be used to construct an effective and efficient prediction model for extreme climate related natural disasters occurrences in the country.

2.0 Literature Review

Over the course of several decades, Zambia has experienced a number of climatic change difficulties. Droughts, seasonal and flash floods, all occur often and high heat and dry periods have been the most severe circumstances lately. Bank [5], observes that droughts and floods, in particular, have become more often, intense, and widespread, posing a danger to food and water security, water quality, energy, and rural populations' long-term survival.

Other studies indicate that Zambia, unusually, is subjected to two climatic extremes at the same time. While residents in the southwest are normally concerned about the country's unpredictable rainfall pattern, those in the northeast are anxious about flash floods [6]. During the 2019/2020 rainy season, the country received above-average rainfall, resulting in floods in some districts in the Southern, Eastern and Lusaka Provinces, while dry spells were experienced in some parts of the Western Province. Over the previous three decades, floods and droughts have cost Zambia more than $13.8 billion in catastrophic losses [7]. Zambia's agricultural output is expected to plummet by 30% by 2080 under the current climate change scenario [1]. Moreover, according to a government study on the economic impacts of climate change conducted in 2011, a loss of roughly USD 5 billion in GDP is expected over a 10-to-20-year timeframe [9]. The World Bank Knowledge Portal offers some insight into the frequency, impact and occurrence of common natural hazards in Zambia as summarized in Figure 2.1.

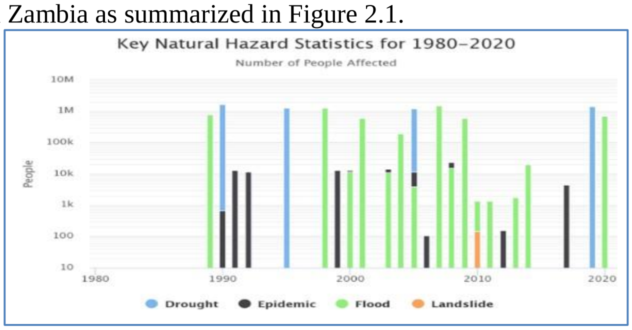


Figure 2.1: Key Natural Disasters occurring in Zambia

Climate variability and unpredictability continue to be a major structural risk to Zambia's long-term growth, affecting important sectors like agriculture and energy and underscoring the need for climate-smart solutions to be included into the country's long-term economic plan.

Early Warning Systems (EWS) are critical components of climate change adaptation, with the goal of preventing or minimizing damage caused by risks. EWS are complex systems that attempt to lessen the consequences of natural catastrophes by providing timely and relevant information in a systematic manner. EWS can help poor nations like Zambia become more resilient to natural disasters and climate-related threats if effectively implemented.

Even though various studies and initiatives have been conducted on disaster vulnerability assessment and prediction models, the literature on climate change induced disaster prediction and EWS for Zambia is scarce. Further, some studies indicate that there is no systematic EWS in place for natural disasters and risks in Zambia [8].

However, a number of efforts have been made to create and strengthen EWSs in the country. These efforts include the dissemination of flood and drought early warnings, including the deployment of community-based EWS in three districts: Chipata, Gwembe, and Sesheke by UNDP [9] and the installation of disaster EWS in Kasaya village and Mbeta Island in the Kazungula and Sioma Districts of the Southern and Western Provinces, respectively, by the ITU in collaboration with ZICTA [10]. In addition, the 21st Century World Bank modernization project, which aims to enhance early warning and disaster management systems, was launched by ZMD and its cooperating partners. The initiative, which began in 2014 and has been ongoing ever since, aims to reduce disaster risk caused by climate change and climatic variability concerns as well as the requirement to lessen operational, budgetary, and human limitations [11].

Despite the highlighted efforts, the nation has continued to suffer significant losses from natural catastrophes since these systems are unable to predict disasters with sufficient accuracy and in a timely manner. These systems mainly depend on weather forecasting that in turn is used to predict the possibility of a hazard occurring not necessarily predicting the actual disasters. Studies have shown that short-term weather forecasts alone are insufficient to safeguard populations during disasters. Recognizing the hazard sooner and implementing the necessary systemic steps are necessary for surviving natural catastrophes.

Machine learning (ML), a branch of Artificial Intelligence (AI) that uses intuitive training to recognize patterns in datasets as part of an algorithmic and heuristic approach, has gained appeal in recent years as an adaptable instrument for technological growth and has been advocated as a useful tool for climate change studies [4].

ML algorithms have also been used to predict severe climate change occurrences and threats. In Zambia, Mzyece et al. [12], researchers from the University of Zambia, sought to apply ML to anticipate seasonal rainfall. Khan et al. [13] investigated the possibility of constructing drought prediction models for Pakistan utilising ML techniques. Similarly, In Ecuador, Muñoz et al. [14] compared ML techniques for powering flood EWS and Moon et al. [15] proposed using ML approaches to create an efficient EWS for extremely short-term high rainfall. Kuradusenge et al. [16] forecasted rainfall-induced landslides using ML Models in Rwanda's Ngororero District. In other recent studies in Nigeria and India [17] [18], Ogunjo et al. and Bhimala et al. used ML techniques to predict future COVID-19 cases based on previous infections, using weather parameters such as temperature, and humidity data.

The most frequently employed ML algorithms in the highlighted studies are LR, RF, KNN, NB, DT, and SVM, which are eventually selected for investigation in this study. Choosing a single ML method is quite challenging, as a result all models are assessed by comparing their accuracy on training and test sets before choosing a single model.

3.0 Methodology

3.1. Theoretical Framework

To create a Machine Learning system, it takes more than just choosing a model, training it, and using it on fresh data. Machine learning projects may be organized with the use of frameworks. The *Cross-Industry Standard Process for Data Mining (CRISP-DM)*, is one such framework. The CRISP-DM proposes a six phases methodology for the machine learning process: *Business understanding, Data understanding, Data preparation, Modeling, Evaluation, and Deployment* as illustrated in Figure 3.1 below.

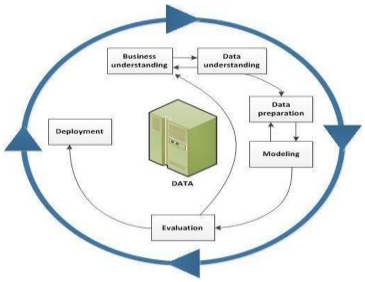


Figure 3.1: CRISP-DM Model (Courtesy IBM, 2021)

This project comprises a model for predicting natural disasters based on patterns and linkages found in past natural disaster records and historical meteorological data. Based on the CRISP-DM framework, the process used included the seven principal steps that are followed in ML model building; *data collection, data preprocessing, model selection, model training, performance evaluation, prediction and model deployment*.

3.2. Data Collection

The historical climate/weather data was obtained from the Zambia Meteorological Department (ZMD) while the historical disaster data was obtained online with the assistance and guidance of the Disaster Management and Mitigation Unit (DMMU). Various online sources also were explored to augment the data provided by the two primary sources. These included but not limited to: (a) *the World Bank Climate Knowledge Portal; (b) the EM-DAT International Disasters Database; (c) the Humanitarian Data Exchange; (d) Our World in Data; and (e) the IRI/LDEO Climate Data Library.*

The Data obtained from ZMD was for the period from July, 1980 to February, 2022 recorded by 40 meteorological stations spread across the country. The key natural disasters that occurred during the corresponding period (1980-2022) were mostly obtained from the World Bank Climate Knowledge Portal and the EM-DAT International Disasters Database.

3.3. Data Preparation and Preprocessing

The obtained data is analyzed, processed and merged to form one meaningful dataset using various tools including Power Query in Excel.

The recorded variables of climate/weather data that constitute numerical features are *temperature (minimum and maximum), humidity, evaporation, pressure, wind speed and rainfall* while the qualitative attributes are *Date* and *Location.* The key recorded natural disasters are *floods, droughts, epidemics, insect infestations* and *landslides*. Insect infestations and landslides were excluded from the final dataset because there were very few recorded instances and their impact was insignificant. The *"Disaster Type"* attribute designate a multi-class target variable with four classes namely; *'Drought', 'Flood',* '*Epidemic* 'and '*No Disaster* 'to represent periods where no disasters were recorded. The final dataset consists of 31509 instances with 10 attributes.

The Visual Studio Code (VS Code) IDE in the Anaconda environment is used to pre-process the dataset and build the model using Python Version 3.9. The Jupyter Notebook and Python Extensions in VS Code IDE are used to write and run the Codes. Some pre-processing techniques such as, *identifying and inputing missing values and outliers, encoding categorical features, Splitting the training and test datasets* and *feature scaling* were undertaken to clean and structure the dataset.

3.4. Model Building

3.4.1 Model Training and Testing

The goal of this study is to train and evaluate different ML models and pick the best for forecasting disasters in Zambia. To compare and find the best feasible model, six (6) classification algorithms are chosen, namely: *Logistic Regression [LR], Random Forest [RF], K-Nearest Neighbor [KNN], Gaussian Naive Bayes [GNB], Decision Tree [DT], and Support Vector Machine [SVM].*

The ML algorithms were trained and tested to evaluate the models. *Eighty percent* (25207 instances) of the dataset is utilised for training, while the remaining *twenty percent* (6302 instances) is used for testing and validation.

3.4.2 Model Evaluation

The Confusion Matrix, Classification Reports, Overfitting/Underfitting tests, Cross Validation scores and AUC-ROC (Area Under the Curve - Receiver Operating Characteristics) plots are used in the evaluation process.

The *Confusion Matrix* is a tool for evaluating classification model performance. It is a classification framework that serves as the foundation for several classification measures. The *True Positives (TP), False Positives (FP), False Negatives (FN),* and *True Negatives (TN)* are the metrics used to assess the performance of the model. Further, the *Classification Report*, is a performance evaluation metric which is used to show the *precision, recall, F1 Score,* and *support score* of the trained classification model. It also displays the *accuracy* of the model.

Overfitting or *underfitting* of the data is the cause of poor model performance. Overfitting refers to a model that models the training data too well while Underfitting refers to a model that can neither model the training data nor generalise to new data. When training and testing data accuracy scores are comparable and nearly equal, there is no underfitting or overfitting and the model generalises effectively to new data.

Also employed in model evaluation is *Cross Validation*, the process of evaluating the correctness of a ML model accuracy when fed with new data. Before deploying the model, Cross Validation score will show whether the model is correct with fresh data if it is close or equal to its accuracy.

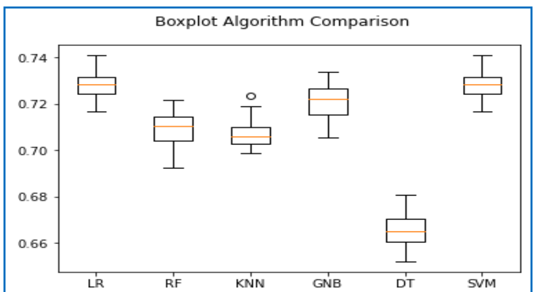
To show a model's performance between *sensitivity* and *specificity*, an AUC-ROC plot is utilised. An AUC-ROC plot, assist to determine how effectively a model can differentiate between classes. The AUC-ROC curve and the other aforementioned metrics are used to evaluate the models, which aids in the justification of results.

4.0 Results and Findings

4.1 Comparison of Models for Selection of Best Performing Al- gorithm

To ensure fair comparison of ML algorithms considered in this study, each algorithm is evaluated in the same way on the same data. Table 4.1 below presents the comparison of the accuracy and cross validation scores. Table 4.2 is a summary of the comparison of training and testing scores to determine whether or not the model is underfitting or overfitting. Further, Figures 4.1 and 4.2 are visualization plots of accuracy and cross validation scores comparison, respectively. Meanwhile, the ROC AUC micro-average scores are 0.85 for LR, KNN and GNB while RF, DT and SVM scored 0.87, 0.78 and 0.86, respectively. Micro-average is preferred in a multi-class classification system if it is suspected that there may be a class imbalance. This would be explained in the sections that follow.

Table 4.1: Comparison of Accuracy and Cross Validation Scores



Model	Accuracy Score	Cross Validation Score	Standard Deviation
LR	0.730720	0.728687	(0.006755)
RF	0.709616	0.709605	(0.008419)
KNN	0.711679	0.707621	(0.007682)
GNB	0.723421	0.720991	(0.009245)
DT	0.666455	0.665291	(0.008366)
SVM	0.730720	0.728687	(0.006755)

Table 4.2: Checking Underfitting and Overfitting of the models

Model	Training Score	Testing Score	Underfitting/ Overfitting
LR	0.728686	0.730720	No
RF	0.900622	0.710568	Yes
KNN	0.758638	0.692796	Yes
GNB	0.721426	0.723421	No
DT	0.900623	0.669311	Yes
SVM	0.728686	0.730720	No

Figure 4.1: Comparison of Model Accuracy

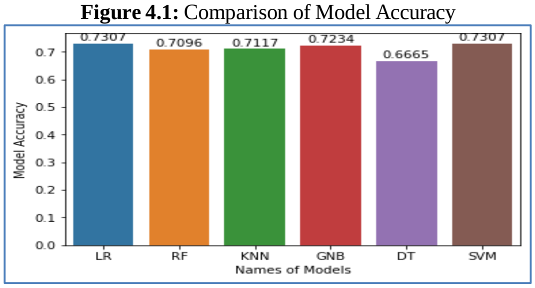


Figure 4.2: Comparison of Cross Validation Scores

4.2 Model Selection

Based on the results above, the algorithms for LR and SVM both outperformed the others. But since LR requires less time to compute than SVM, it was chosen to create the final model. SVM models are successful in large dimensional spaces and often have greater classification accuracy, but their drawback is that they take longer to compute than LR models.

4.3 Resolving Imbalanced Classes Using SMOTE

It is observed that the dataset has a class imbalance, which affects how well the chosen model performs. When tested for predictions, the majority of predictions are slanted in favour of the class *"No Disaster*." The observation is an indication that other classes are in the minority are ignored which results in poor performance of the model.

The lack of sufficient data contributes to the problem of class imbalance and the default behaviour of the ML model is to overclassify the majority class when there is a class imbalance. Although the accuracy of 73% of the proposed model may be good, the resultant model is biased and more likely to misclassify members of the minority class.

In order to make the number of instances in the minority classes more closely match the number of examples in the majority class, oversampling is the practice of synthesising fresh examples of the minority classes. The *Synthetic Minority Oversampling Technique (SMOTE)* is the most used method for creating new samples. This method aids in overcoming the overfitting issue brought on by random oversampling [19]. Before SMOTE is applied, the number of instances in each class is checked and confirmed and the results are listed below:

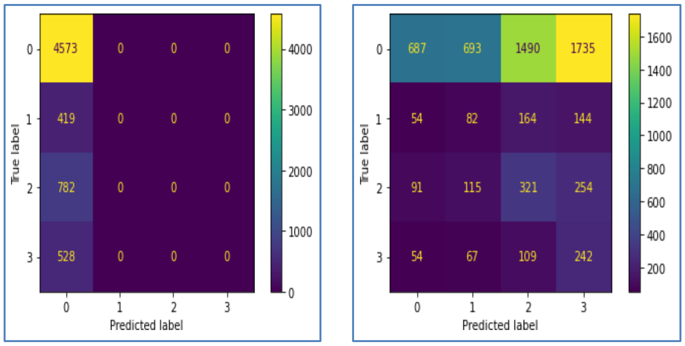
Class="No Disaster", n=22973 (72.9%)
Class="Drought", n=2182 (6.9%)
Class="Flood", n=2562 (8.1%)
Class="Epidemic", n=3792 (12.0%)

The dataset is then oversampled using SMOTE. It oversamples all classes by default so that each class has an equal number of instances to the class with the most instances. The outcomes of using SMOTE are shown below:

Class="No Disaster", n=18378 (25.000%)
Class="Drought", n=18378 (25.000%)
Class="Flood", n=18378 (25.000%)
Class="Epidemic", n=18378 (25.000%)

Using the resampled dataset, the LR model is then retrained.

The transformed Confusion Matrix, which is shown in Figure 4.3 below, demonstrates how effectively the model trained on synthetic cases generalises. The classifier does a good job of recognizing the minority classes. When tested, the model was able to predict *'No Disaster', 'Flood'* and *'Epidemic'* classes when fed with random input values. However, the model could not produce any result for the 'Drought' class probably because the classified number of 82 instances of True Positives for the class is still too low compared to the other classes.



Before SMOTE After SMOTE
Figure 4.3: Transformed Confusion Matrix after SMOTE

4.4 Deployment of the Model

The process for developing a ML model is complete with model deployment as illustrated in Figure 4.4.

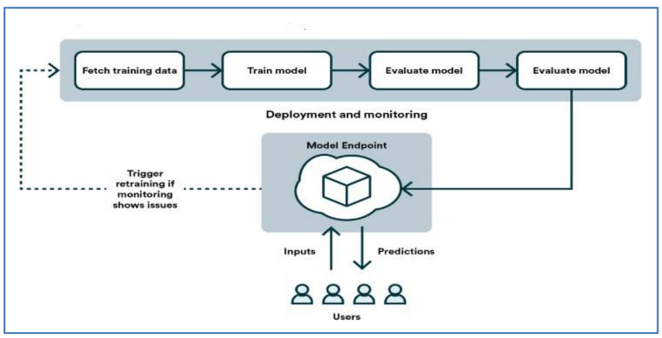


Figure 4.4: Model Building and Deployment Pipeline

The three options of model deployment are all taken into consideration. These are deployment for batch prediction, deployment as a web service and deployment on Android devices. Due to their simplicity and low cost, the second and third options were selected and used in this project.
The web service application was developed using the flask framework [20] and the Android application was developed by utilising Java in Android Studio to produce the application [21]. Both applications are to be deployed to platforms such as Heroku and Google Play where the model would be accessible to those concerned.

5.0 Discussion and Analysis

The primary source of class imbalances is when classifiers trained on uneven training sets have a prediction bias, which causes poor performance in the minority classes [38]. In the case of this study, the historical disasters dataset was not regularly recorded. In certain instances, the recorded weather parameters could point out on the possibility of a disaster occurring, such as a flood or drought, in a particular period, but no disaster is recorded for that year.
This study further established that DMMU does not host any database for disasters. The Early Warning Section of the Unit when approached, indicated that it also relies on online sources for historical disasters data and provided a number of websites where such data could be obtained. The data obtained from the online sources was also inadequate and contributed to the imbalance of classes that affected the prediction power of the model. Despite using SMOTE to solve the problem, research shows that the method has its own flaws. Models that are trained on oversampled datasets when applied to real-world situations may fail because classifiers trained on these samples while being told that they represent minorities, may provide predictions that are off when the models are applied to actual data [22].
In addition, the analysis of the dataset during the preparation and pre-processing stages, also revealed that apart from the issue of unbalanced classes, the obtained historical climate data from ZMD are also incomplete, full of missing values and outliers, and to a large extent of low quality and integrity because it also contained extreme and wrong values.
Outlier detection is a crucial problem in data mining. The most popular thresholding strategies are based on statistics like interquartile range, median absolute deviation, and standard deviation around the mean. The inclusion of outliers in the calculation of these statistics might generate severe distortions and their usage therefore may not be accurate [23]. In this project, almost all variables have a high number of outliers that could lead to inaccurate predictions after these data points are removed.
Similarly, the issue of missing values also has an unfavourable impact on the outcomes of analyses, particularly when it results in inaccurate parameter estimations [24]. In the case of this study, it can be shown that the number of missing values is quite high for all parameters. The statistical mean was used to replace these missing values. Various studies indicate that data imputation using mean can influence errors of the prediction if the percentages of missing values is high and the size of the dataset is small [25]. The model performance of this project is likely to be affected by this pre-process.

In climate studies and applications, such as analysis of climate change trends and modeling their effects and solutions on various socioeconomic activities, good quality climate data are crucial. However, Zambia still grapples with the challenge of maintaining a meteorological and disasters database with reliable quality data that can be analyzed to draw inferences and create solutions.
It is evident that ZMD and DMMU are still facing challenges in maintaining reliable databases based on the quality of the data obtained. ZMD, for instance, still largely relies on using the traditional manual weather measuring methods and instruments for most of the weather observations. Weather stations are still scarce and in addition, their locations are unequal, with the majority being in urban areas. This is despite ZMD having embarked on the modernization program in 2014 that involved the installation of Automated Weather Observation Stations (AWOS) [11]. Given this situation, the suggested model's performance suffers even after the dataset has been cleaned and organized using various ML techniques.

6.0 Conclusion and Recommendation

The study was conducted by comparing six ML algorithms to get the best predictive model. The model making process was carried out by the seven principal stages of ML model building process of data collection, data preprocessing, model selection, model training, performance evaluation, prediction and model deployment which are based on the CRISP-DM methodology. The results show that the LR and SVM algorithms has the best accuracy in predicting with a precision value of 73% but LR is chosen for model deployment because it has less computational time compared to SVM.

However, model performance is affected due to imbalanced classes, outliers, and missing values in the dataset that influence biased predictions. ML techniques are employed to remove outliers and impute missing values. Further, SMOTE is utilized to address the problem of imbalanced classes. Nonetheless, these techniques have their own flaws and the model is expected not perform perfectly in a real-world data environment.

The challenge of acquiring historical weather and disasters data for analysis is likely to persist if ZMD and DMMU do not improve their operations in collecting and maintaining reliable databases. It is therefore recommended that efforts to improve the performance of the two institutions is enhanced probably through automation of their systems so that data could be collected and stored efficiently and with reduced errors.

Future work will focus on further study of the proposed model to improve its performance by imploring other ML algorithms especially in the realm of Deep Learning.

References

[1] J. S. Phiri, “Adaptation of Zambian Agriculture to Climate Change- A Comprehensive Review of The Utilisation of The Agro-Ecological Regions,” *Climate Change*, p. 41, 2013.

[2] J. Rawlins and F. K. Kalaba, “Adaptation to Climate Change: Opportunities and Challenges from Zambia,” in *African Handbook of Climate Change Adaptation*, N. Oguge, D. Ayal, L. Adeleke, and I. da Silva, Eds. Cham: Springer International Publishing, 2021, pp. 2025–2044. doi: 10.1007/978-3-030-45106-6_167.

[3] Ministry of lands, Natural Resources and Environmental Protection, “National Policy on Climate Change: Zambia,” 2016. https://www.mlnr.gov.zm/?wpfb_dl=74 (accessed Nov. 19, 2021).

[4] D. Rolnick *et al.*, “Tackling Climate Change with Machine Learning,” 2019. [Online]. Available: www.climateinformatics.org

[5] A. D. Bank, “Zambia - National Climate Change Profile,” *African Development Bank - Building today, a better Africa tomorrow*, Sep. 16, 2019. https://www.afdb.org/en/documents/zambia-national-climate-change-profile (accessed Nov. 10, 2021).

[6] Friday Phiri, “Climate Change: A Tale of Weather Extremes with Mixed Fortunes for Zambia - Zambia,” *ReliefWeb*, 2020. https://reliefweb.int/report/zambia/climate-change-tale-weather-extremes-mixed-fortunes-zambia (accessed Nov. 12, 2021).

[7] J. Mwitwa, “Zambia National Drought Plan,” p. 112, 2018.

[8] C. C. Makondo, K. Chola, and B. Moonga, “Climate Change Adaptation and Vulnerability: A Case of Rain Dependent Small-Holder Farmers in Selected Districts in Zambia,” *American Journal of Climate Change*, vol. 03, no. 04, pp. 388–403, 2014, doi: 10.4236/AJCC.2014.34034.

[9] Nelson Gapare, “Strengthening climate information and early warning systems in Eastern and Southern Africa for climate resilient development and adaptation to climate change – Zambia (PIMS 5091) Final Report June 2019,” 2019. https://pims.undp.org/attachments/5091/213789/1726196/1740680/TE_5091_Final_04_06_19.pdf (accessed Nov. 24, 2021).

[10] Onder Cetinkaya, “Natural Disaster Early Warning System in zambia,” 2018. https://www.itu.int/en/ITU-D/Emergency-Telecommunications/Documents/2018/Zambia/Zambia_EWS_Report.pdf (accessed Nov. 11, 2021).

[11] O. Mudenda and E. Nkonde, “Lessons from The Modernization of National Meteorological and Hydrological Services (NMHS) – A Case Study of The Zambia Meteorological Department.,” p. 10, 2018.

[12] L. Mzyece, M. Nyirenda, M. Kabemba, and G. Chibawe, “Forecasting Seasonal Rainfall in Zambia – An Artificial Neural Network Approach,” *Zambia ICT Journal*, vol. 2, p. 16, Jun. 2018, doi: 10.33260/zictjournal.v2i1.46.

[13] N. Khan, D. A. Sachindra, S. Shahid, K. Ahmed, M. S. Shiru, and N. Nawaz, “Prediction of droughts over Pakistan using machine learning algorithms,” *Advances in Water Resources*, vol. 139, p. 103562, 2020, doi: 10.1016/J.ADVWATRES.2020.103562.

[14] P. Munoz, J. Orellana-Alvear, J. Bendix, and R. Célleri, “Comparison of Machine Learning Techniques Powering Flood Early Warning Systems. Application to a catchment located in the Tropical Andes of Ecuador.,” Copernicus Meetings, EGU2020-4243, Mar. 2020. doi: 10.5194/egusphere-egu2020-4243.

[15] S.-H. Moon, Y.-H. Kim, Y. H. Lee, and B.-R. Moon, “Application of machine learning to an early warning system for very short-term heavy rainfall,” *Journal of Hydrology*, vol. 568, pp. 1042–1054, Jan. 2019, doi: 10.1016/j.jhydrol.2018.11.060.

[16] M. Kuradusenge, S. Kumaran, and M. Zennaro, “Rainfall-Induced Landslide Prediction Using Machine Learning Models: The Case of Ngororero District, Rwanda,” 2020, doi: 10.3390/ijerph17114147.

[17] S. T. Ogunjo, I. A. Fuwape, and A. B. Rabiu, “Predicting COVID-19 Cases From Atmospheric Parameters Using Machine Learning Approach,” *GeoHealth*, vol. 6, no. 4, p. e2021GH000509, 2022, doi: 10.1029/2021GH000509.

[18] K. R. Bhimala, G. K. Patra, R. Mopuri, and S. R. Mutheneni, “Prediction of COVID-19 cases using the weather integrated deep learning approach for India,” *Transboundary*

and Emerging Diseases, vol. 69, no. 3, pp. 1349–1363, 2022, doi: 10.1111/tbed.14102.

[19] Swastik Satpathy, "SMOTE | Overcoming Class Imbal- ance Problem Using SMOTE," *Analytics Vidhya*, Oct. 06, 2020. https://www.analyticsvidhya.com/blog/2020/10/overcoming- class-imbalance-using-smote-techniques/ (accessed Jun. 29, 2022).

[20] T. Ghanoum, "Complete Guide on Model Deployment with Flask and Heroku," *Medium*, Jan. 01, 2022. https://to-wardsdatascience.com/complete-guide-on-model-deployment- with-flask-and-heroku-98c87554a6b9 (accessed Jun. 23, 2022).

[21] R. Agrawal, "How to Deploy Machine Learning(ML) Model on Android," *Analytics Vidhya*, Nov. 17, 2021. https://www.analyticsvidhya.com/blog/2021/11/how-to-de- ploy-machine-learningml-model-on-android/ (accessed Nov. 12, 2022).

[22] A. S. Tarawneh, A. B. Hassanat, G. A. Altarawneh, and A. Almuhaimeed, "Stop Oversampling for Class Imbalance Learn- ing: A Review," *IEEE Access*, vol. 10, pp. 47643–47660, 2022, doi: 10.1109/ACCESS.2022.3169512.

[23] Jiawei Yang, Susanto Rahardja, and Pasi Fränti, "Outlier detection | Proceedings of the International Conference on Ar- tificial Intelligence, Information Processing and Cloud Compu- ting," *ACM Other conferences*, 2019. https://dl.acm.org/doi/abs/10.1145/3371425.3371427 (ac- cessed Jun. 30, 2022).

[24] N. Z. Abidin, A. Ritahani, and N. A., "Performance Anal- ysis of Machine Learning Algorithms for Missing Value Impu- tation," *ijacsa*, vol. 9, no. 6, 2018, doi: 10.14569/IJACSA.2018.090660.

[25] W. Badr, "6 Different Ways to Compensate for Missing Data (Data Imputation with examples)," *Medium*, Jan. 12, 2019. https://towardsdatascience.com/6-different-ways-to-compen- sate-for-missing-values-data-imputation-with-examples- 6022d9ca0779 (accessed Nov. 10, 2022).

[26] Godana, T., & Mukwena, R. M. (2004). Intergovernmental Relations and Fiscal Decentralisation in Namibia. Governance in Southern Africa and Beyond: Experiences of institutional and public policy Reform in Developing Countries. Windhoek: Gamsberg MacMillan, 85-110.

[27] Mukwena, R. (2000). Crisis, Adjustment and Social Change in Zambia: The Case of Professionals. Lusaka, Zambia.

[28] Mukwena, R. M. (2004). The Role of Decentralization in Reducing Regional Inequalities in Namibia. Regional Development Dialogue, 25, 108-128.

[29] Mukwena, R. (2014). Decentralisation, democracy and development: The case of Zambia. 50 Years of Local Government in Zambia: Treasuring the Past, Reflecting the Present, Shaping the Future, 22-70.

[30] Mukwena, R. M. (1999). Can local government performance be measured? Lessons from Zambia. Africanus, 29(1), 45-58.

[31] Mukwena, R. M. (1998). The role of local councils in rural development: A study of Gwembe and Kalomo District Councils, Zambia, 1981-1995. The University of Manchester (United Kingdom).