ZKG INTERNATIONAL

ISSN: 2366-1313

AN INNOVATIVE FLOOD PREDICTION USING MACHINE LEARNING MODELS APPROACH

D.LAVANYA[1], M.SANTHAN GNANENDRA[2], N. PRAVEEN KUMAR[3], M.REVATHI SAI[4], K. NANDINI[5]

[1]Assistant Professor, CSE,Chalapathi Institute of Technology,Guntur, India
[2]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India
[3]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India
[4]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India
[5]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India

ABSTRACT: Machine Learning (ML) models for flood prediction can be beneficial for flood alerts and flood reduction or prevention. To that end, machine-learning (ML) techniques have gained popularity due to their low computational requirements and reliance mostly on observational data. This study aimed to create a machine learning model that can predict floods in Kebbi state based on historical rainfall dataset of thirty-three years (33), so that it can be used in other Nigerian states with high flood risk. In this article, the Accuracy, Recall, and Receiver Operating Characteristics (ROC) scores of three machine learning algorithms, namely Decision Tree, Logistic Regression, and Support Vector Classification (SVR), were evaluated and compared. Logistic Regression, when compared with the other two algorithms, gives more accurate results and provides high performance accuracy and recall. In addition, the Decision Tree outperformed the Support Vector Classifier. Decision Tree performed reasonably well due to its above-average accuracy and below-average recall scores. We discovered that Support Vector Classification performed poorly with a small size of dataset, with a recall score of 0, below average accuracy score and a distinctly average roc score.

ISSN: 2366-1313

1. INTRODUCTION

Among the natural disasters, floods are the most destructive, causing massive damage to human life, infrastructure, agriculture, and the socioeconomic system. Governments, therefore, are under pressure to develop reliable and accurate maps of flood risk areas and further plan for sustainable flood risk management focusing on prevention, protection, and preparedness. Flood prediction models are of significant importance for hazard assessment and extreme event management. Robust and accurate prediction contribute highly to water recourse management strategies, policy suggestions and analysis, and further evacuation modeling. Thus, the importance of advanced systems for short-term and long-term prediction for flood and other hydrological events is strongly emphasized to alleviate damage .However, the prediction of flood lead time and occurrence location is fundamentally complex due to the dynamic nature of climate condition. Therefore, today's major flood prediction models are mainly data-specific and involve various simplified assumptions. Thus, to mimic the complex mathematical expressions of physical processes and basin behavior, such models benefit from specific techniques e.g.,event-driven, empirical black box, lumped and distributed, stochastic, deterministic, continuous, and hybrids. Physically based models were long used to predict hydrological events, such as storm [, rainfall/runoff shallow water condition hydraulic models of flow and further global circulation phenomena including the coupled effects of atmosphere, ocean, and floods Although physical models showed great capabilities for predicting a diverse range of flooding scenarios, they oftenrequire various types of hydro-geo morphological monitoring datasets, requiring intensive computation, whichprohibits short-term prediction Furthermore, as stated in Reference the development of physically based models often requires in-depth knowledge and expertise regarding hydrological parameters,reported to be highly challenging. Moreover, numerous studies suggest that there is a gap in short-term prediction capability ofphysical model major improvements in physically based models of flood were recently reported through the hybridization of models as well as advanced flowsimulations In addition to numerical and physical models, data-driven models also have a long tradition in flood modeling, which recently gained more popularity. Data-driven methods of prediction assimilate the measured climate indices and hydro-meteorological parameters to provide better insight. Among them, statistical models of autoregressive moving average (ARMA) multiple linear regression (MLR) and autoregressive integrated moving average (ARIMA) are the most common flood frequency analysis (FFA) methods for modeling flood prediction. FFA was among the early statistical methods for predicting floods Regional flood frequency analyses (RFFA) more advanced versions, were reported to be more efficient when compared to physical models considering computation cost and generalization. Assuming floods as stochastic processes, they can be predicted using certain probability distributions from historical stream flow data For instance, the climatology average method (CLIM) empirical orthogonal function (EOF) multiple linear regressions (MLR), quintile

ISSN: 2366-1313

regression techniques (QRT) and Bayesian forecasting models are widely used for predicting major floods. However, they were reported to be unsuitable for short-term prediction, and, in this context, they need major improvement due to the lack of accuracy, complexity of the usage, computation cost, and robustness of the method. Furthermore, for reliable long term prediction, at least, a decade of data from measurement gauges should be analyzed for a meaningful forecast In the absence of such a dataset, however, FFA can be done using hydrologic models of RFFA, e.g., MISBA and Sacramento as reliable empirical methods with regional applications, where stream flow measurements are unavailable. In this context, distributed numerical models are used as an attractive solution. Nonetheless, they do not provide quantitative flood predictions, and their forecast skill level is “only moderate” and they lack accuracy . The drawbacks of the physically based and statistical models mentioned above encourage the usage of advanced data-

ISSN: 2366-1313

driven models, e.g., machine learning (ML). A further reason for the popularity of such models is that they can numerically formulate the flood nonlinearity, solely based on historical data without requiring knowledge about the underlying physical processes. Data-driven prediction models using ML are promising tools as they are quicker to develop with minimal inputs. ML is a field of artificial intelligence (AI) used to induce regularities and patterns, providing easier implementation with low computation cost, as well as fast training, validation, testing, and evaluation, with high performance compared to physical models, and relatively less complexity. If the data is scarce or does not cover varieties of the task, their learning falls short, and hence, they cannot perform well when they are put into work. Therefore, using robust data enrichment is essential through, e.g., implementing a distribution function of sums of weights invariance assessments to retain the group characteristics or recovering the missing variables using causally dependent coefficients The second aspect is the capability of each ML algorithm, which may vary across different types of tasks. This can also be called a "generalization problem", which indicates how well the trained system can predict cases it was not trained for, i.e., whether it can predict beyond the range of the training dataset. For example, some algorithms may perform well for short-term predictions, but not for long-term predictions. These characteristics of the algorithms need to be clarified with respect to the type and amount of available training data, and the type of prediction task, e.g., water level and stream flow. In this review, we look into examples of the use of various ML algorithms for various types of tasks. At the abstract level, we decided to divide the target tasks into short-term and long-term prediction. We then reviewed ML applications for flood-related tasks, where we structured ML methods as single methods and hybrid methods. Hybrid methods are those that combine more than one ML method. Here, we should note that this paper surveys ML models used for predictions of floods on sites where rain gauges or intelligent sensing systems used. Our goal was to survey prediction models with various lead times to floods at a particular site. From this perspective, spatial flood prediction was not involved in this study, as we did not study prediction models used to estimate/identify the location of floods. In fact, we were concerned only with the lead time for an identified site.

2. LITERATURE REVIEW

The rainfall forecasting is prevailing as a popular research in the scientific areas in the modern world of technology and innovation; as it has a huge impact on just the human life but the economies and the living beings as a whole. Rainfall prediction with several Neural Networks has been analyzed previously and the researchers are still trying hard to achieve the more perfect and accurate results in the field of rainfall prediction (Biswas, et al., 2016). The prediction of seasonal rainfall on monthly basis by using the surface data to form annual prediction is also essential for the agricultural activities and therefore the production and supervision of the agriculture and crops. It could be done by recognizing the variations in the supply of

ISSN: 2366-1313

moisture in the air. The case of African region illustrates that how this succeeded and how West Africa advantaged from the rainfall prediction in managing their agricultural Similarly, the short-term stream flow forecasting for the rainfall is also reliable and bias-free. But they are not much effective in predicting the flood and post-processing of rainfall prediction. An approach called raw numerical weather prediction (NWP) was introduced in 2013, where the approach focused on the Bayesian joint probability model to formulate prediction data.

3. EXISTING SYSTEM

The existing system deals with improving the quality of dataset which is being used. Data mining approach helps to find the hidden pattern, which will help to predict the rainfall correctly. This approach takes all the parameters, which affect the rainfall such as climate, wind speed etc. and predict the future rainfall. Customized, integrated and modified data mining technique is used to predict rainfall.

4. PROPOSED SYSTEM

We have proposed ML based rainfall prediction and forecasting system to efficiently predict the rainfall and to do forecasting for upcoming years. It provides the better accuracy comparing to the existing approach. It consumes less time for huge amount of data.

5. IMPLEMENTATION

Preprocessing of Data:

A data set is a collection of data. Dataset contains only the matrix of numerical data. We use the data set which contains the rainfall data from 1901 to 2015 for different regions across the country. For Data pre-processing we use the most commonly used min-max normalization method to convert all rainfall intensity values to a number between 0 and 100

Modeling:

Here We are applied Various Machine learning algorithms applied. Such as

Random forest

LogisticRegression

DecisionTree

KNN

SVC

K-Nearest Neighbors (KNN)

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not **"**learn**"** until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data
- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables **(**non-metric variables**)**
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

Naïve Bayes

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to thepresence (or absence) of any other feature .

ISSN: 2366-1313

Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervisedlearning techniques. Various reasons have been advanced in the literature. In this tutorial, wehighlight an explanation based on the representation bias. The naive bayes classifier is a linearclassifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vectormachine). The difference lies on the method of estimating the parameters of the classifier (the learning bias).While the Naive Bayes classifier is widely used in the research world, it is not widespread amongpractitioners which want to obtain usable results. On the one hand, the researchers found especiallyit is very easy to program and implement it, its parameters are easy to estimate, learning is very fasteven on very large databases, its accuracy is reasonably good in comparison to the otherapproaches.

Random Forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance. The first algorithm for random decision forests was created Random forests are frequently used as "black box" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

SVM

A discriminant function that can correctly predict labels for newly acquired instances**.** Unlike generative machine learning approaches, which require computations ofconditional probability distributions, a discriminant classification function takes a data point *x* and assignsit to one of the different classes that are a part of the classification task**.** Less powerful than generativeapproaches, which are mostly used when prediction involves outlier detection, discriminant approachesrequire fewer computational resources and less training data, especially for a multidimensional featurespace and when only posterior probabilities are needed**.** From a geometric perspective, learning a classifieris equivalent to finding the equation for a multidimensional surface that best separates the different classesin the feature space**.**SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyper plane parameter—in contrast to*genetic algorithms* **(***GAs***)** or*perceptrons*, both of which are widely used for classification in machine learning**.** For perceptron's, solutions are highly dependent on the initialization and termination criteria**.** For a specific kernel that transforms the data from the input space to the feature space, trainingreturns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GAclassifier models are different each time

ISSN: 2366-1313

training is initialized**.** The aim of GAs and perceptrons is only tominimize error during training, which will translate into several hyperplanes' meeting this requirement**.**

6. RESULT

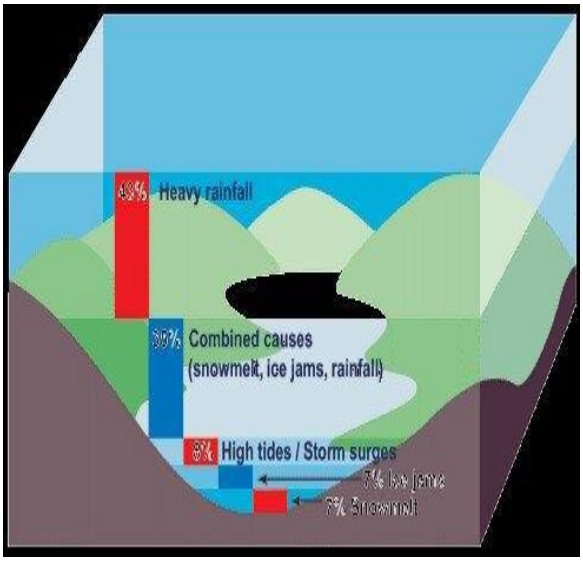

FIG: Flood Prdictio By Using Improved Machine Learning

7. CONCLUSION

This project proposed a novel combination of LR and SGD as a voting classifier for emotion recognition by classifying tweets as happy or unhappy. Our experiments showed that one can improve the performance of models by recognizing patterns efficiently and through effective averaging combination of models. Experiments are conducted to test various machine learning models thatare; GBM, LR, DT, andVC(LR-SGD). This study also employed two feature representation techniques TF and TF-IDF. The results showed that all models performed well on tweet dataset but our proposed voting classifier VC(LR-SGD) outperforms by using both TF and TF-IDF among all. Proposed model achieves the highest results using TF-IDF with 78% Accuracy, 80% Recall and 81% F1-score.

REFERENCES

1. Abhishek, K., Kumar, A., Ranjan, R., & Kumar, S. (2012). A Rainfall Prediction Model using Artificial Neural Network. IEEE Control and System Graduate Research Colloquium, 1- 5.
2. Abraham, A. (2005). Artificial Neural Networks. John Wiley and Sons, Ltd. Adler, R. F., Huffman, G. J., Bolvin, D. T., Curtis, S., &Nelkin, E. J. (2000). Tropical Rainfall Distributions Determined Using TRMM Combined with Other Satellite and Rain Gauge Information. American Metrological Society, 1-9.
3. Agnihotri, G., & Panda, J. (2014). Comparison of Rainfall from Ordinary and Automatic Rain Gauges in Karnataka. Mausam, 65(4), 1-8. Ahn, J. (2017). Analysis of a neural network model for building energy hybrid controls for inbetween season. Architecture of Complexity, (pp. 1-5).
4. Akingbaso, E. Y. (2014). Land Use - Cover Change Assessment Framework: Famagusta North Cyprus. Approval of the Institute of Graduate Studies and Research-EMU.
5. Alpers, W., &Melsheimer, C. (2004). Rainfall. SAR Marine User Manual, US Dept. of Commerce,
6. NOAA. Amoo, O. T., &Dzwairo, B. (2016). Trend analysis and artificial neural networks forecasting for rainfall trends. Environmental Economics, 7(4), 1-10.
7. bbc. (n.d.). Weather. Retrieved December11, 2017, from cypnet.co,uk: http://www.cypnet.co.uk/ncyprus/main/weather/index.html Beard, J. S. (1962). Rainfall interception by grass. South African Forestry Journal, 42(1), 1-7.
8. Biswas, S. K., Marbaniang, L., Purkayastha, B., Chakraborty, M., Singh, H. R., &Bordoloi, M. (2016). Rainfall forecasting by relevant attributes using artificial neural networks - a comparative study. International JOurnal of Big Data Intelligence, 3(2), 1-10. 73

ISSN: 2366-1313

9. Blyth, A. M., Bennett, L. J., & Collier, C. G. (2015). High-resolution observations of precipitation from cumulonimbus clouds. Meteorological applications, 22, 1-12.
10. Carvalho, J. R., Assad, E. D., Oliveira, A. F., & Pinto, H. S. (2014). Annual maximum daily rainfall trends in the Midwest, southeast and southern Brazil in the last 71 years. Weather and Climate Extremes, 5(6), 1-7.
11. Cheng, M., & Qi, Y. (2001). Frontal Rainfall-Rate Distribution and Some Conclusions on the Threshold Method. Journal of applied meteorology, 41, 1-8.
12. Collier, C. G. (2003). On the formation of stratiform and convective cloud. Weather,58, 1-4.
13. Darji, M. P., Dabhi, V. K., &Prajapati, H. B. (2015). Rainfall Forecasting Using Neural Network: A Survey. International Conferences on Advances in Computer Engineering and Applications (pp. 1-5). Ghaziabad, India: IEEE.
14. Dayan, U., Ziv, B., Margalit, A., Morin, E., & Sharon, D. (2001). A severe autumn storm over the middle-east: synoptic and mesoscale convection analysis. Theoretical and Applied Climatology, 69(1-2), 1-8.