A NOVEL APPROACH FOR PREDICTION OF RAINFALL USING LOGISTIC REGRESSION TO COMPUTE ACCURACY AND ERROR RATE AND COMPARING WITH RANDOM FOREST ALGORITHM

K. Lakshmi Poojitha[1], R. Dhanalakshmi[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: Heavy rains can create a number of tragedies in the rain. Forecasting is essential. The prediction allows people to take preventative steps, and it must produce the desired outcome.
Materials and methods: In the proposed work categorizing is performed by adopting a pattern size of n =10 with the g-power value of 80% and datasets are collected from various web sources with recent study findings and threshold 0.05%, confidence interval 95% mean and standard deviation was iterated 20 times to obtain data in the Random Forest algorithm. The Random Forest algorithm and implementation will be used in it. For the implementation, an additional test could be used**.**
Results: It shows a low accuracy of random forest (87.01%) in contrast with the logistic regression by the set of rules (95.25%). There is a statistically insignificant difference between the observed agencies with sizable 0.075 ($p>0.05$)
Conclusion: Prediction is a category of rainfall prediction that indicates that the selection tree seems to generate higher accuracy than the random forest.

Keywords: Machine learning, Rainfall Prediction, Linear Regression, Novel Classification, Decision Tree, Random Forest, Logistic Regression

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and technical Science, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.
[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

1. Introduction

The research work is the capacity of locating a few algorithms in different values that is primarily based at the prediction of rainfall and some Precipitation figuring out into numerous sorts of rainfall inside the crucial in light of the fact that weighty and unpredictable precipitation could have many effects like annihilation of harvests and homesteads, harm of property so a superior expecting model is essential for an early word which could limit dangers to life and assets and moreover handling the farming ranches in better manner (Hin 2014). This novel type expectation essentially helps ranchers and moreover water assets can be used proficiently (Chen and Georgakakos 2015). Precipitation expectation is an issue in device mastering missions and the consequences have to be unique. There are several equipment gadgets for expecting precipitation by means of utilizing the climate conditions like machine,linear, temperature, mugginess, and pressure. (Badarpur.2003) The software of research work is Precipitation dedication is large for some catchment board packages, mainly for flood note frameworks (Zhang et al. 2019). The profundity of precipitation and its conveyance inside the fleeting and spatial measurements is based upon several elements, like a couple of regression pressure, temperature, and wind velocity and heading (Heavy Rainfall Prediction the use of Gini 2014).

In a few study papers of rainfall prediction, the wide variety of papers published in the studies gate is 824, and the number of papers published in google scholar is 23,800 (G. 2020). This implementation will help farmers in figuring out which crops to reap primarily based on crop seasons (Lee and Georgakakos 1991).The effects of this paper suggests that the xg boost set of rules and linear regression like system and linear is giving possibility for rainfall prediction. In this paper, we see many using deep algorithms to classify the predictions. According to paper effects, Implements a model that forecasts weather situations which includes rainfall, fog, thunderstorms, and cyclones, permitting people to take precautions and novel classification (Badarpura et al. 2020). Various techniques produce numerous correctnesses so choose the right calculation and version it as indicated by using **t**he prerequisites. (Pujari 2001). The above survey concluded that the algorithm did not have accuracy prediction; here we find the accuracy whilst locating that it shows many problems in the algorithms like novel,gadget,linear regression ("Heavy Rainfall Prediction Using Gini Index in Decision Tree" 2019)Our team has extensive knowledge and research experience that has translated into high quality publications(Pandiyan et al. 2022; Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022)

The above literature survey concluded that the algorithm did not have accuracy prediction; here we find the accuracy while finding that it shows many issues in the algorithms (Lee and Georgakakos 1991). It means that a small change in the data can lead to a large change in the structure of the decision tree. They are often like linear and multiple linear regression, relatively inaccurate and linear regression ((Selamat, Hakeem Selamat, and Rais 2015)). The existing approaches have poor accuracy. There are other predictions that perform better with similar data. The aim of machine learning in this study is to implement an efficient method for rainfall prediction by using logistic regression and random forest (Mahmood 2017).

2. Materials and Methods

The research was performed in the Image Processing Lab, Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS. Basically it is considered that two groups of classifiers are used namely XGboost and random forest algorithms, which is used to classify the sentimeThe above literature survey concluded that the algorithm did not have accuracy prediction; here we find the accuracy while finding that it shows many issues in the algorithms (Lee and Georgakakos 1991). It means that a small change in the data can lead to a large change in the structure of the decision tree. They are often like linear and multiple linear regression, relatively inaccurate and linear regression ((Selamat, Hakeem Selamat, and Rais 2015)). The existing approaches have poor accuracy. There are other predictions that perform better with similar data. The aim of machine learning in this study is to implement an efficient method for rainfall prediction by using logistic regression (Mahmood 2017).

Logistic Regression

It is used in statistical software to understand the relationship between the dependent variable and one or more independent variables by estimating probabilities using a logistic regression this type of analysis can help you predict the likelihood of an event happening.Logistic regression is a useful analysis method for novel classification problems, where you are trying to determine if a new sample fits best into a category. Logistic regression does not face these strict assumptions and is much more robust when these assumptions are not yet,making

its application appropriate in many situations and some researchers prefer logistic regression because it is similar to multiple regression.

Pseudo Code for Logistic Regression

Step 1: Start
Step 2: Designed for survival analysis of algorithms
Step 3: It has predictiction of group membership in different ways
Step 4: Predict the research result of the given algorithm
Step 5: Check the accuracy values of logistic regression
Step 6: Weather it is possible in some cases which is under logistic regression
Step 7: To analyze and check the logistic regression accuracy
Step 8: End

Random Forest

Random forest is a machine learning technique that's used to solve regression and novel classification problems.it utilizes ensemble learning ,which is a technique that combines many classifiers to provide solutions to complex problems. A random forest is a supervised machine learning algorithm that is constructed from decision tree algorithms. It is applied in various industries such as banking and ecommerce to predict behavior and outcomes.It improves the accuracy of machine learning algorithms and it reduces the overfitting of datasets and increases precision and also generates predictions without requiring many configurations in some packages and provides an effective way of handling missing data.

Pseudo Code for Random Forest Algorithm

Step 1: Start the process
Step 2: It performs in different process and algorithms which is based on pseudocode
Step 3: Check the accuracy values in different process
Step 4: Random forest classifier and fit in data
From sklearn ensemble import Random forest classifier
Step 5: It is widely used in novel classification and regression problems
Step 6: Choose the no of algorithms performed in the pseudocode
Step 7: End

Statistical Analysis

The analysis was done using IBM SPSS version 21. It is a statistical software tool used for data analysis. For both proposed and existing algorithms 10 iterations were done with a maximum of 20 samples and for each iteration the predicted accuracy was noted for analyzing accuracy. The value obtained from the iterations of the Independent Sample T-test was performed. The independent data sets are targets, hashtag, flag. The Dependent values are sent through. A detailed analysis has been done on these values for finding accuracy rate in rainfall prediction. The independent variables of the data are accuracy (Koning and Smith 2017), Standard deviation, and standard mean error and dependent variables in the machine multiple regression data are in the dataset and the parameters of the data file. The independent sample T-test analysis is carried out in this research work.

3. Results

The dataset is provided by a novel approach for prediction of rainfall using logistic regression and random forest to compute accuracy and error rate. Table 1 shows the weather conditions in different ways to check the accuracy. Table 2 shows the number of iterations the accuracy of Logistic regression varies random forest low for Depression prediction with a mean value is 95.25 %, Std. Deviation is 1.70368. Thus the model is able to work efficiently to predict the rainfall in different algorithms like machine learning,novel classification,and linear regression. The mean difference ,standard deviation difference and significant difference of Decision tree based rainfall detection and logistic regression based rainfall detection is tabulated in Table 3,which shows there is a significant difference between the two groups since $P<0.05$ with an independent sample T-Test. The linear regression dependent variables are minimum and maximum temperatures shows the rainfall prediction with the help of independent variables are date, location. The statistical analysis of two independent groups shows that the logistic regression has higher accuracy mean (95.25%) compared to random forest Fig. 1.

4. Discussion

Based on the above look at its miles found that the Logistic regression algorithm has higher accuracy of 95.25% than Random forest area which has 87.01% in prediction of rainfall. There is a statistically insignificant difference between the observed agencies with sizable 0.075 ($p>0.05$) by acting impartial pattern t-assessments.

Similar findings algorithms are very extensively efficient and accurate in predicting high accuracy of rainfall prediction in comparison to the prevailing set of rules, (Campilho and Kamel 2008) this is Decision tree (Stark and Bowyer 1996). The dataset containing a large quantity of

pictures is given as input into each of the algorithms and the accuracy price of (Stark and Bowyer 1996; Jiang et al. 2018) prediction is received for the present and the proposed algorithms like system,linear regression. These values acquired are used for evaluation and assessment for efficiency linear and machine (Selamat, Hakeem Selamat, and Rais 2015). (Campilho and Kamel 2008). The contrary findings have applied the detection of the usage of assessment of pooling operations at the home windows that are not having tremendous improvement on over non-overlapping pooling home windows (Wilkowski, Stefańczyk, and Kasprzak 2020). This function without delay compares them into rainfall duties like novel category,gadget,linear regression (Kumar, Upadhyay, and Senthil Kumar 2020).

In this paper logistic is given high-quality accuracy while comparing it to random woodland (Suk and Bhandarkar 2012). The model developed in this work does a good job of classifying whether defaults will occur. It has a positive limitation impact on financial business on risk management. And we can establish which attributes are essential elements in determining the default rate using the models feature importance analysis. In future the mean error rate can be approved by using different methods.

5. Conclusion

In this paper a compiled list of the most current developments in rainfall prediction and detection approaches, as well as their drawbacks and future research directions. There are several rainfall detection strategies, datasets, and comparisons explored. We have also developed a taxonomy of the techniques and identified the established results.The results show that the proposed rainfall prediction compares the outcomes of the Logistic Regression with Random Forest. Among two models, Logistic Regression has obtained better accuracy 95.25% than the Random Forest algorithm which has an accuracy of 87.014%. The precision of the rainfall prediction has been significantly increased.

Declarations

Conflict of Interests

No conflict of interest.

Authors Contribution

Data collection, data analysis, and manuscript writing were all done by author KLP.Author RD was involved in

the Action process, Data verification and validation, and Critical review of manuscript.

Acknowledgements

The authors would like to thank Saveetha School of Engineering and Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the infrastructure necessary to successfully complete this research.

Funding: We would like to thank the following organizations for providing financial assistance that allowed us to complete the study.

1 . SNEW Technologies, Chennai.
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

6. References

Baraneedharan, P., Sethumathavan Vadivel, C. A. Anil, S. Beer Mohamed, and Saravanan Rajendran. 2022. "Advances in Preparation, Mechanism and Applications of Various Carbon Materials in Environmental Applications: A Review." *Chemosphere*. https://doi.org/10.1016/j.chemosphere.2022.134596.

Deena, Santhana Raj, A. S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2022. "Enhanced Biogas Production from Food Waste and Activated Sludge Using Advanced Techniques – A Review." *Bioresource Technology*. https://doi.org/10.1016/j.biortech.2022.127234.

Karpagam, M., R. Beaulah Jeyavathana, Sathiya Kumar Chinnappan, K. V. Kanimozhi, and M. Sambath. 2022. "A Novel Face Recognition Model for Fighting against Human Trafficking in Surveillance Videos and Rescuing Victims." *Soft Computing*. https://doi.org/10.1007/s00500-022-06931-1.

Koning, Mark, and Chris Smith. 2017. *Decision Trees and Random Forests: A Visual Introduction for Beginners*. Independently Published.

Kumar, P. Ganesh, P. Ganesh Kumar, Rajendran Prabakaran, D. Sakthivadivel, P. Somasundaram, V. S. Vigneswaran, and Sung Chul Kim. 2022. "Ultrasonication Time Optimization for Multi-Walled Carbon Nanotube Based Therminol-55 Nanofluid: An Experimental Investigation." *Journal of Thermal Analysis and Calorimetry*. https://doi.org/10.1007/s10973-022-11298-4.

Lee, Tim H., and Konstantine P. Georgakakos.

1991. *A Stochastic-Dynamical Model for Short-Term Quantitative Rainfall Prediction*.

Mahmood, Arshad. 2017. *Rainfall Prediction in Australia: Clusterwise Linear Regression Approach*.

Nagarajan, Karthik, Arul Rajagopalan, S. Angalaeswari, L. Natrayan, and Wubishet Degife Mammo. 2022. "Combined Economic Emission Dispatch of Microgrid with the Incorporation of Renewable Energy Sources Using Improved Mayfly Optimization Algorithm." *Computational Intelligence and Neuroscience* 2022 (April): 6461690.

Nagaraju, V., B. R. Tapas Bapu, P. Bhuvaneswari, R. Anita, P. G. Kuppusamy, and S. Usha. 2022. "Role of Silicon Carbide Nanoparticle on Electromagnetic Interference Shielding Behavior of Carbon Fibre Epoxy Nanocomposites in 3-18GHz Frequency Bands." *Silicon*. https://doi.org/10.1007/s12633-022-01825-1.

Pandiyan, P., R. Sitharthan, S. Saravanan, Natarajan Prabaharan, M. Ramji Tiwari, T. Chinnadurai, T. Yuvaraj, and K. R. Devabalaji. 2022. "A Comprehensive Review of the Prospects for Rural Electrification Using Stand-Alone and Hybrid Energy Technologies." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102155.

Selamat, M. Hakeem, M. Hakeem Selamat, and Helmi Md Rais. 2015. "Image Face Recognition Using Hybrid Multiclass SVM (HM-SVM)." *2015 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*. https://doi.org/10.1109/ic3ina.2015.7377765.

Venu, Harish, Ibham Veza, Lokesh Selvam, Prabhu Appavu, V. Dhana Raju, Lingesan Subramani, and Jayashri N. Nair. 2022. "Analysis of Particle Size Diameter (PSD), Mass Fraction Burnt (MFB) and Particulate Number (PN) Emissions in a Diesel Engine Powered by Diesel/biodiesel/n-Amyl Alcohol Blends." *Energy*. https://doi.org/10.1016/j.energy.2022.123806.

Whangchai, Niwooti, Daovieng Yaibouathong, Pattranan Junluthin, Deepanraj Balakrishnan, Yuwalee Unpaprom, Rameshprabu Ramaraj, and Tipsukhon Pimpimol. 2022. "Effect of Biogas Sludge Meal Supplement in Feed on Growth Performance Molting Period and Production Cost of Giant Freshwater Prawn Culture." *Chemosphere* 301 (August): 134638.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Advances in the Application of Immobilized Enzyme for the Remediation of Hazardous Pollutant: A Review." *Chemosphere* 299 (July): 134390.

Tables and Figures

Table 1. Comparison of accuracy was achieved during the evaluation of Logistic Regression and Random Forest models for novel classification with different iteration

S.NO	Logistic Regression	Random Forest
1	87.10	95.23
2	85.49	93.22
3	83.35	91.24
4	82.25	94.35
5	84.32	92.33
6	82.25	91.35
7	86.33	94.45
8	83.21	95.34
9	86.40	93.33
10	87.23	94.22

Table 2. Group Statistics of Logistic regression with random forest by grouping the iterations with Sample size 10, Mean = 95.2520, Standard Derivation = 1.70368, Standard Error Mean = .76191. Descriptive Independent Sample Test of Accuracy and is applied for the dataset in SPSS. Here it specifies Equal variances with and without assuming a T-Test Score of two groups with each sample size of 10

Accuracy	Algorithm	N	Mean	Standard Deviation	Standard Error Mean
1	Logistic Regression	10	95.2520	1.70368	.76191
2	Random Forest	10	87.0160	1.46877	.65685

Table 3. Independent Sample Test of Accuracy for Logistic Regression and Random Forest

Accuracy	Levene's Test for Equality of Variances		t-test for Equality of Means							
	F	Sig.	t	df	One-sided p	Two-sided p	Mean difference	std.error difference	95% Confidence Interval of the Difference	
									Lower	Upper
Equal variances assumed	3.576	.075	6.478	18	<.000	<.001	7.67300	1.18441	5.18465	10.16135
Equal variances not assumed			6.478	17.360	<.001	<.001	7.67300	1.18441	5.17806	10.16794

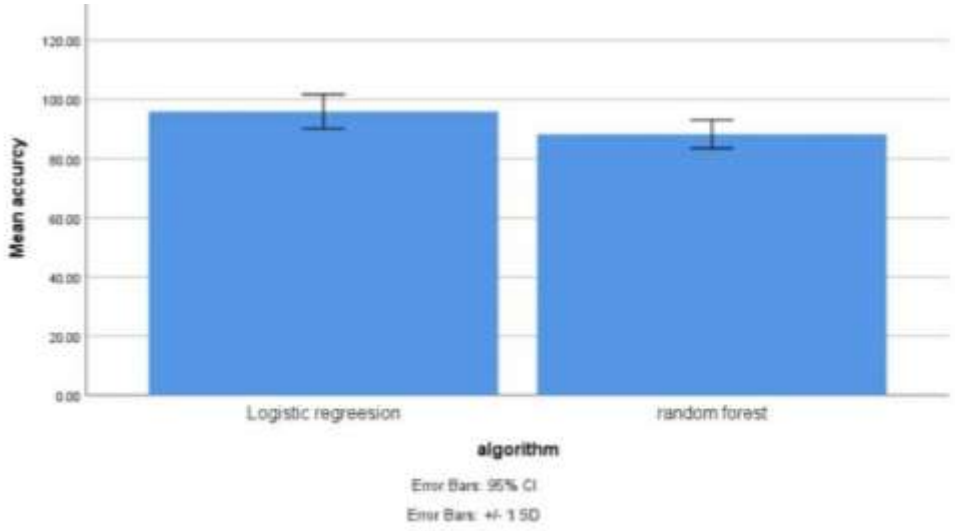

Fig. 1. Comparison of Logistic Regression over Random Forest in terms of mean accuracy. It explores that the mean accuracy is slightly better than the Random forest and the standard deviation is moderately improved compared to the Random forest. Graphical representation of the bar graph is plotted using groupid as X-axis DS vs XGB Y-Axis displaying the error bars with a mean accuracy of detection +/- 1 SD.