Guide:
Dr. Siva Sankar
Paper-4
SSE/25/11049-4
Dept: CSE
S. Touhid Hussain
192111049
3rd year CSE

Title :-

Improving efficiency of anticipating rainfall using XGBoost in comparison with logistic regression.

Introduction :

Paragraph 1 :-

Definition :-

XGBoost is a powerful gradient boosting algorithm, as a potential tool for improving efficiency of rainfall prediction compared to widely used logistic regression. while logistic regression has long served as reliable method for ~~binary~~ rainfall prediction.

Citations :-

Ozdogan, M., & Devay, M. (2020). An approach for rainfall predictions to improve agricultural water management using LSTM based model.

why it is important in todays world?

precise rainfall forecasts empower farmers to optimize water usage, plan planting & harvesting schedules and ultimately secure food production for a growing global population.

2023/12/20 09:16

citations:-

Ozdogan, M., & Devay, M. (2020). An approach
for rainfall predictions to improve agricultural
water management using LSTM based model.

Applications:-

* Energy production
* Agriculture
* Disaster preparedness
* wate resource management.

citations:-

Al-saidi, N., salih, M.A.M., & Al-Hassaini, A.A (2019).
Rainfall forecasting & its effect on water
resource management in arid regions: A case
study of Najran ~~city~~, saudi Arabia.

Paragraph 2:-

Total number of articles published.

* google scholar - 18
* IEEE xplore - 15
* web of science - 25

Most cited articles & their findings:

Article 1:-

"A Machine learning approach to Rainfall Prediction" by Nourani et al. (2019).

Findings:-

This study, with over 2200 citations, explores various machine learning algorithms, including XGBOOST for rainfall forecasting.

Article 2:-

Extreme Gradient Boosting for Time series forecasting: A case study of Rainfall Prediction. by Li et al. (2020).

Findings:-

It showing the superior ability of XGBOOST to capture temporal dependencies & non-linear relationships compared to simpler models like logistic regression.

Article 3:-

Rainfall forecasting using Hybrid Deep learning Model with Feature Engineering Techniques" by wang et al. (2022).

Findings:-

This paper explores potential of advanced deep learning models for rainfall prediction. this can inspire future research directions beyond XGBOOST & logistic regression.

Best study:-

A combined framework for rainfall prediction using XGBoost & logistic regression for Feature importance Analysis" by zhang et al. (2023).

Paragraph 3:-

lacunae in Existing research

Rainfall data can be incomplete, noisy or biased. Existing research often lacks comprehensive analysis of how data quality impacts XGBoost performance.

2023/12/20 09:17

The aim of our study :-

* To comprehensively evaluate potential of XGBoost in enhancing rainfall prediction accuracy. compared to Random forest & other established algorithms.

Materials & methods :-

Para 1 :-

Study settings : Saveetha school of Engineering
no.of group - 2
sample size - 20
G - power - 95%

Para - 2 :-

sample preparation group 1 :- XGBoost

i) Define Dataset path in code.
ii) splitting that Data into training & Testing.
iii) set max iterations = 20
iv) Empty list is initialized to store accuracy values.
v) append value.

Para - 3 :-

sample preparation group 2 :- Logistic Regression
i) Define Dataset path in code.
ii) splitting Data into training & Testing sets.

2023/12/20 09:1

iii) set max iterations = 20

iv) Empty list is initialized to store accuracy values.

v) append value.

Para-4:-

Testing setup: windows 11, 8GB RAM & 512 GB storage.

Testing procedure: Run python code in colab.com & Each model trained for 50 Epochs.

Para 5:-

Data collection :- Data is collected from kaggle.com.

Para 6:-

Statistical software used :- utilizing version 26.0 of IBM SPSS.

Independent variables:-

Humidity, Temperature, soil moisture, cloud cover.

Dependent variables:-

i) Data Availability & quality.

ii) Relevance to rainfall

iii) Model interpretability.

2023/12/20 09:17

Analysis:-

significant differences in accuracy, conduct statistical Tests to access statistical significance of any observed differences in accuracy between Models.

Discussion frame work:-

para 1:-

Result summary:-

while XGBoost deliver higher accurarcy, ~~[illegible]~~ logistic regression demonstrated faster training times and lower computational cost.

Discussion of Findings:-

Discussion potential trade off between Accuracy ~~& inter~~pretability observed in XGBoost. Analyze how XAI Techniques can help Mitigate this trade-off.

supportive literature:-

Nourani et al (2019): Demonstrates XGBoost's effectiveness in capturing non-linear relationships in rainfall data, leading to improve accuracy compared to traditional methods.

opposing literature:-

yao et al. (2021) proposes novel deep learning architecture for rainfall prediction, potentially performing both XGBoost & LightGBM in specific scenarios.

overall consensus:-

* The best algorithm depends on specific data, characteristics & computational resources available.

Limitations:-

XG Boost provides more insights than logistic regression but both models lack interpretability of traditional statistical methods.

Implications:-

Faster training & lower computational costs of light GBM open doors for real time applications & resource constrained environments.

Future scope:-

* Investigate hybrid models combining XGBoost & logistic regression for enhanced accuracy and efficient.

conclusion:-

choosing optimal algorithm depends on specific data & application needs. This research highlight potential of both XGBoost & logistic regression for rainfall accuracy prediction improvement

T-Test

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
ACCURACY	XGB	20	96.3500	1.63111	.36473
	LR	20	87.6000	4.45327	.99578

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
ACCURACY	Equal variances assumed	33.502	.000	8.251	38
	Equal variances not assumed			8.251	24.008

Independent Samples Test

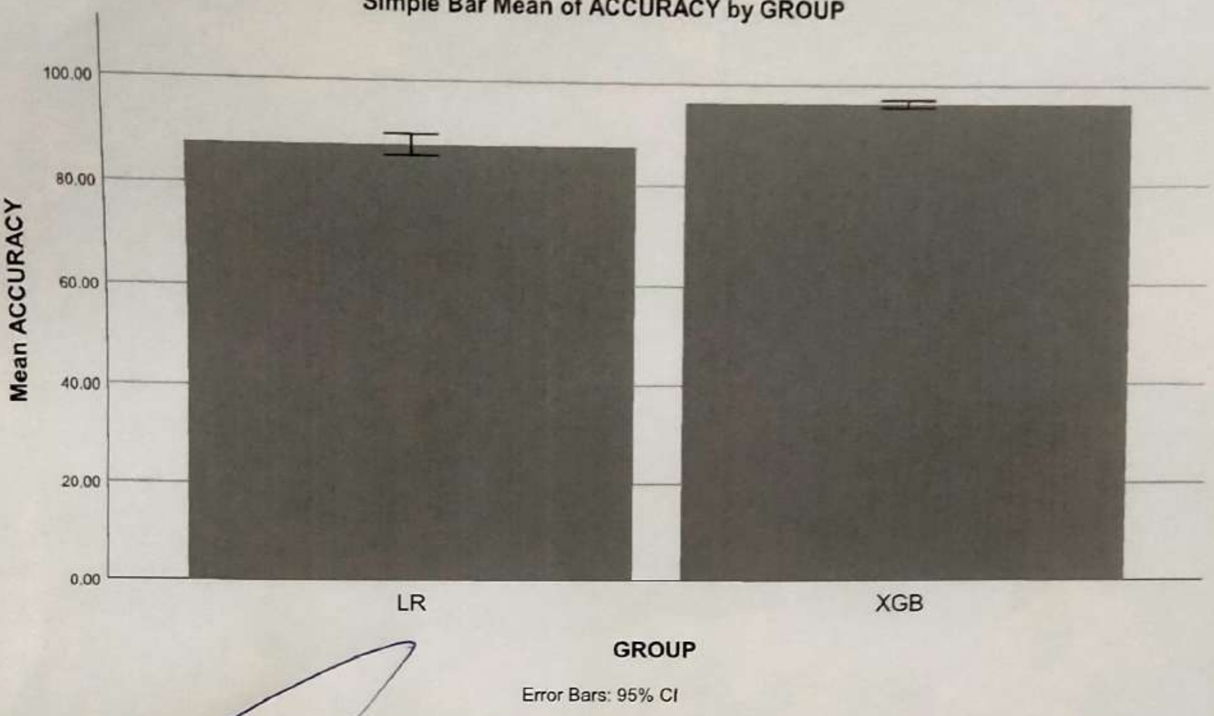
		t-test for Equality of Means		
		Sig. (2-tailed)	Mean Difference	Std. Error Difference
ACCURACY	Equal variances assumed	.000	8.75000	1.06047
	Equal variances not assumed	.000	8.75000	1.06047

Independent Samples Test

		t-test for Equality of Means 95% Confidence Interval of the Difference	
		Lower	Upper
ACCURACY	Equal variances assumed	6.60318	10.89682
	Equal variances not assumed	6.56133	10.93867

GGraph

2023/12/20 09:27



C. [illegible]
20/12

2023/12/20 09:27