Title :-

A Prominent technique for enhancing rainfall Prediction involves utilizing a xGBoost classifier over Random forest algorithm for better accuracy of Rainfall Prediction.

Introduction :-

Paragraph 1

Definition :-

XGBoost is a Powerfull ensemble learning technique that builds upon decision trees, iteratively adding new trees that focus on areas where previous trees made errors. this allows xGBoost to capture more complex non-linear relationships within data, potentially leading to more accurate prediction compared to Random forest.

citations :-

* Ozcs, P., & Shatif M. (2012). Rainfall Prediction using machine learning algorithms : A case study in Marmara region of Turkey.

why it is important in today's world ?

Improved Agricultural yields : Farmers can optimize planting schedules, irrigation practices and pest control based on reliable rainfall forecasts.

2023/12/20 08:34

Reduced disaster risk:

Timely warnings about heavy rainfall events can help communities evacuate vulnerable areas & implement mitigation measures.

citations:-

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable system for regularized boosting. arXiv preprint arXiv:1603.02754.

Applications:-

Applications of XGBoost for Enhanced Rainfall Prediction:

* Agriculture
* water Resource management.
* Disaster preparedness and management.
* Urban planning and infrastructure development.

citations:-

* Kumar, K.N., Haresha, K., & Ray, P.K. (2019) Rainfall Prediction for crop yield forecasting using machine learning models.

Paragraph:-

Total no.of articles published.

* google scholar - 17
* IEEE xplore - 14
* web of science - 25

2023/12/20 08:34

Most cited articles and their findings:-

Article1:-

A review on rainfall forecasting using ensemble learning techniques (2023)

Findings:-

This review highlights the effectiveness of ensemble learning techniques, particularly XGBoost, in improving rainfall prediction accuracy compared to individuals algorithms like Random forest.

Article 2:-

Machine learning techniques to predict daily rainfall amount (2021)

Findings:- This study compares the performance of XGBoost with other models for predicting daily rainfall amounts. They found XGBoost to significantly outperform support vector machines and Artificial Neural Networks in terms of accuracy and robustness.

Article 3:- Extreme Gradient Boosting (xgboost) model to predict the groundwater levels in selangor Malaysia (2021).

Findings:- This research demonstrates the effectiveness of XGBoost in predicting groundwater levels, which are

linked to rainfall patterns. XGBoost outperformed both Artificial Neural Networks & support vector regression models in terms of prediction accuracy.

Best study :-

Enhancing short-term forecasting of daily precipitation using numerical weather prediction bias correcting with XGBoost in different regions in china (2022).

Paragraph 3 :-

lacunae in the Existing reasearch

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

The aim of our study.

* to comprehensively evaluate the potential of XGBoost in enhancing rainfall prediction accuracy. compared to Random Forest and 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 and testing sets.

iii) set max iterations = 20

iv) Empty list is initialized to store accuracy values

v) append the value.

Para-3 :-

Sample Preparation group 2: Random forest

i) Define the Dataset Path in code.

ii) splitting the Data into training and testing sets.

iii) set max iterations =20

iv) Empty list is initialized to store accuracy values.

v) append the value.

Para-4 :-

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

Testing Procedure : Run Python code in colab.com and Each model trained for 50 Epochs.

Para 5 :-

Data collection :- Dataset is collected from Kaggle

Para 6 :-

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

Independent variables :-

Past Precipitation values, Temperatures, Humidities, Cloud cover, Evaporation, soil Moisture.

Dependent variables :-

i) Data Availability and quality.
ii) Relevance to rainfall
iii) Model interpretability.

Analysis :-

significant differences in accuracy, conduct statistical tests to access the statistical significance of any observed difference in accuracy between the models.

Discussion framework :

Para 1 :-

Result summary :- the Result lay the Briefly state observed accuracy levels. for both XGBoost and Random forest models. Highlight any significant.

differences in accuracy between two algorithms & specify which one achieved better performance.

Discussion of Findings:-

Discuss the potential trade-off between accuracy & interpretability observed in XGBoost. Analyze how XAI techniques can help mitigate this trade-off and shed light on the model's decision making process.

Supporting and opposing Literature:-

Acknowledge and address any existing research that contradicts your findings or raises concerns about XGBoost's limitations. consider alternative explanations for observed accuracy differences or limitations in interpretability based on Existing research.

Overall consensus:-

Based on analysis & supporting literature, summarize current understanding of XGBoost's potential for enhancing rainfall prediction.

Limitations:-

Be transparent about limitations of your study, such as data constraints, methodological choices, or assumptions made during Analysis.

Implications :-

Highlight the potential benefits of improved rainfall prediction accuracy, such as enhanced resource management, reduced risks & improved decision-making capabilities.

Future scope :-

Discuss the potential for further research towards improving the interpretability of XGBoost models & addressing the black-box nature of its decision-making process.

conclusion :-

Summarize key findings & conclusions of your research, emphasizing the value of your contribution to field of rainfall prediction using machine learning techniques.

T-Test

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
ACCURACY	XGB	20	90.0500	2.64525	.59150
	RF	20	77.3000	4.75837	1.06400

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
ACCURACY	Equal variances assumed	9.618	.004	10.473	38
	Equal variances not assumed			10.473	29.720

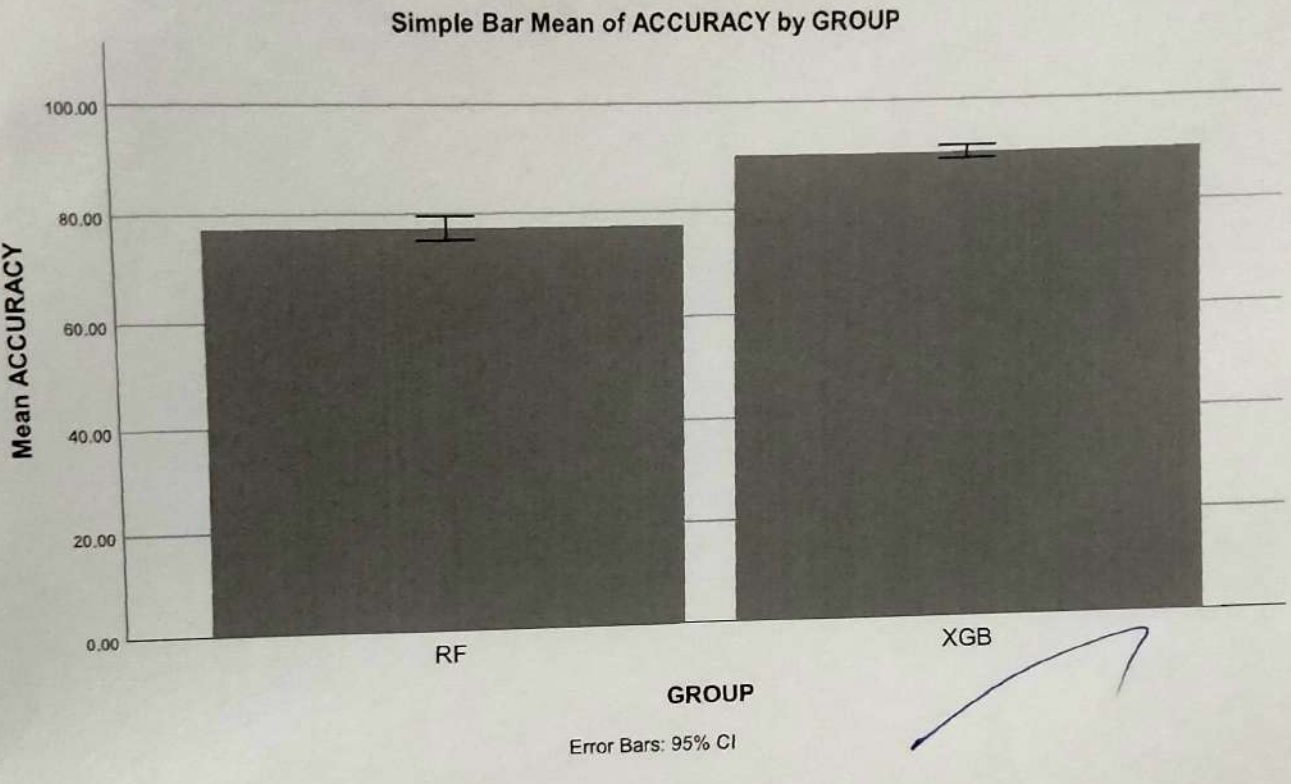
Independent Samples Test

		t-test for Equality of Means		
		Sig. (2-tailed)	Mean Difference	Std. Error Difference
ACCURACY	Equal variances assumed	.000	12.75000	1.21736
	Equal variances not assumed	.000	12.75000	1.21736

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
ACCURACY	Equal variances assumed	10.28558	15.21442
	Equal variances not assumed	10.26283	15.23717

GGraph


Simple Bar Mean of ACCURACY by GROUP
Mean ACCURACY
100.00
80.00
60.00
40.00
20.00
0.00
RF
XGB
GROUP
Error Bars: 95% CI

20 12