SSE/25-11/0479-3
Dept: CSE

Title :-

the foreMOST Technique for enhancing rainfall forecasting using XGBoost algorithm over ~~Light~~ Decision tree GBM classifier for improvent of accuracy.

Introduction

Paragraph 1 :-

Definition :-

XGBoost, a powerful gradient boosting algorithm, to improve accuracy of rainfall forecasting. Traditionally, decision tree have been employed for this task. XGBoost aims to proper rainfall forcasting towards greater Precision & reliability.

citations :-

Aech, V., Ahamed, T., & Nath, T.K. (2015). Rainfall intensity forcast & flood inundation studies for Brahmaputra in India using coupled MM5-WRF mesoscale model.

why it is important in today's world?

precise fore casts empower farmers to optimize irri gation practices, plan planting & harvesting &

2023/12/20 09:14

ultimately enhance crop yields & food security.

citation:

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

Applications :-

* agriculture
* Energy Production
* water resource management.
* Flood management.

citations :-

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

Paragraph 2 :-

Total no. of articles published.

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

most cited articles & their findings :-

Article 1 :-

Deep learning for Rainfall prediction: A comprehensive Review by Fan et al. (2020)

2023/12/20 09:15

Findings :-
this review Paper over 500 citations, Provides comprehesive overview of deep learning techniques applied to rainfall forecasting.

Article 2
Extreme gradient Boosting for Time series forecasting: A case study of Rainfall prediction by Li et al. (2020).

Findings.
It highlights its superior performance over decision trees in capturing temporal dependencies & non-linear relationships within time series data.

Article 3 :-
A machine learning Approach to Rainfall Prediction by Nourani et al. (2019).

Findings:-
this is highly cited over 2200 citations Explores application of various machine learning algorithms for rainfall forecasting.

Para-?

Best study :-

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

Paragraph 3:-

lacunae in Existing research

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

The aim of study:-

* To comprehensively evaluate potential of XG Boost in enhancing rainfall prediction Accuracy compared to Decision tree algorithm.

Materials & Methods :-

Para 1:-

Study settings: saveetha school of Engineering

no.of groups - 2

sample size - 20

G-power - 95 %

2023/12/20 09:15

Para-2:-

sample preparation group 1 : XG Boost.

Sample Preparation group 1 : XG Boost.

i) Define Dataset path in code.

ii) split Data into training & testing sets.

iii) set max iterations = 20

iv) Empty list is initialized to store accuracy values.

v) append value.

Para 3:-

Sample Preparation group 2: Decision tree.

i) Define Dataset Path in code.

ii) split Data into training & Testing sets.

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 & 512GB Storage.

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

Para 5:-

Data collection :- Dataset is collected from kaggle.com

Para 6:-

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

2023/12/20 09:1

Independent variables:-

past precipitation values, Temperature, Humidity Evaporation, soil moisture.

Dependent variables:-

i) Data quality & Availability.

ii) Relevance to rainfall.

iii) model interpretability.

Analysis:-

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

Discussion framework:-

para 1:-

Resolt summary:-

XGBoost is gradient boosting algorithm, leading to increased model complexity compared to decision trees. This complexity likely contributed to higher accuracy but also resulted in:

* longer training times

* low interpretability of models internal workings.

Discussion of Findings:-

Discuss potential trade-off between Accuracy & interpretability observed in XGBoost. Analyze how XAI Techniques can help mitigate this trade-off.

2023/12/20 09:15

Supportive literature:

Li et al.(2020): showcases XGBoost's superior ability to capture temporal dependencies & non-linear relationships within time series data. compared to decision trees for rainfall prediction.

Opposing literature:-

Yao et al. (2021): proposes novel deep learning architectures for rainfall prediction, potentially out performing both XG-Boost & decision tree in specific scenarios.

Overall consensus:-

Decision trees offer advantages in faster training times & easier interpretability but tend to yield lower accuracy.

Limitations:-

→ Limited or noisy data can hinder the performance of both XGBoost & decision trees.

→ XGBoost's higher accuracy comes at cost of lower interpretability compared to decision trees.

Implications:-

→ Improved rainfall prediction accuracy can be benefit agriculture, flood management, water resource management.

→ Trade-off between accuracy & interpretability requires careful consideration for specific applications.

Future scope:-

Develop methods for improving the interpretability of XGBOOST to bridge the gap with Decision trees.

→ Explore incorporating XGBOOST into real-time forcasting systems for near-term predictions.

Conclusion:-

This research highlights potential of XGBoost for enhancing rainfall forecasting accuracy compared to decision trees. Choosing the optimal model depends on specific data & application needs.

2023/12/20 09:16

T-Test

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
ACCURACY	XGB	20	96.8500	1.42441	.31851
	DT	20	81.6000	4.01838	.89854

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
ACCURACY	Equal variances assumed	13.758	.001	15.997	38
	Equal variances not assumed			15.997	23.701

Independent Samples Test

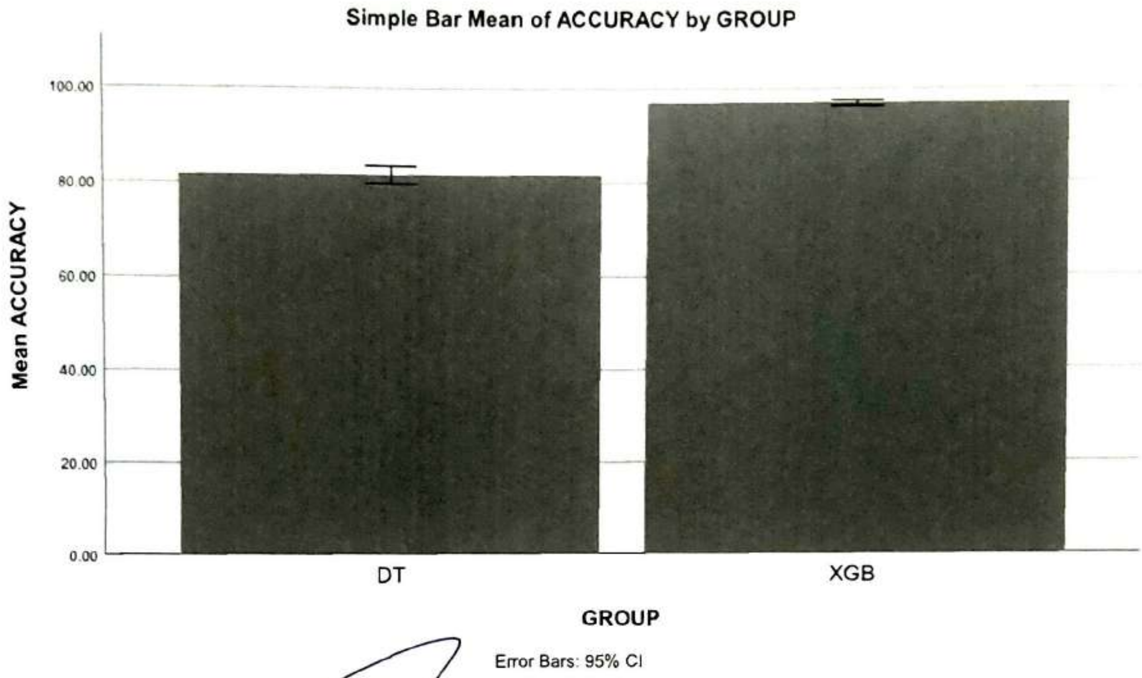
		t-test for Equality of Means		
		Sig. (2-tailed)	Mean Difference	Std. Error Difference
ACCURACY	Equal variances assumed	.000	15.25000	.95332
	Equal variances not assumed	.000	15.25000	.95332

Independent Samples Test

		t-test for Equality of Means 95% Confidence Interval of the Difference	
		Lower	Upper
ACCURACY	Equal variances assumed	13.32011	17.17989
	Equal variances not assumed	13.28113	17.21887

GGraph

2023/12/20 09:16


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