Guide:
Dr. Sivasankar.

Paper-2

SSE/25/11/0479-2
Dept: CSE

S. Towhid Hussain
192111049
3rd year CSE

TITLE :-

An efficient prediction of Rainfall using XGBOOST algorithm over Light GBM classifier for improvement of accuracy.

Introduction

Paragraph 1 :-

Definition :-

By exploring the capabilities of XGBoost, a powerful ~~[illegible]~~ gradient boosting algorithm known for its accuracy & efficiency, and comparing it to Light GBM, the study aims to identify superior approach for rainfall prediction.

citations :-

* Mishra, V., Cherkauer, K.A., & Shakula, M.O. (2020). A vision for global precipitation prediction at hourly time scales.

why it is important in todays world ?

* It impacts everything from food security and disaster preparedness to sustainable water management and economic stability.

2023/12/20 08:31

H
* *

* Efficient allocation of water resources for irrigation, urban water supply. By anticipating periods of drought | heavy rainfall, communities can optimize water usage & prevent shortage or floods.

citations :-

Al-saidi, N., sahil, M.A.M., & Al-Hussaini, A.A. (2012) Rainfall forecasting & its efficient of water resource management in Arid regions: A case study of Najran city. saudi Arabia.

Applications :-

the applications are.

* Agriculture
* Flood management
* water Resource management.
* Energy production.

citations :-

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

Paragraph 2 :-

Total no. of articles published.

* google scholar - 18

2023/12/20 08:31

* IEEE xplore - 15
* web of science - 24

Most cited articles and their findings:-

Article 1 :-

"A Machine learning approach to Rainfall prediction" by Dourani et al. (2019).

Findings :-

this study, with over 2000 citations explore use of various machine learning algorithms, including XGBoost, for rainfall forecasting.

Article 2 :-

Rainfall prediction using Artificial neural Networks and LightGBM by Nhat et al. (2021)

Findings :-

A comparative study" by Nhat etal. (2021), this article over 1000 citations compares ANN & LightGBM for rainfall prediction. reveals strength of LightGBM in high accuracy.

2023/12/20 08:32

Article 3 :-

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

Findings :-

This review paper, with over 500 citations, provides a comprehensive overview of deep learning techniques applied to rainfall forecasting.

Best study :-

~~A case stud~~ Hyperparameters optimization in XGBoost Model for Rainfall Estimation: A case study in Pontinak city.

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.

The aim of study :-

* To comprehensively evaluate potential of XGBoost in enhancing rainfall prediction Accuracy compared to Light GBM & other established algorithms.

2023/12/20 08:33

Materials & methods :-

Para 1 :-

study sittings: saveetha school of Engineering

no. of groups - 2

sample size - 20

G - power - 95 %.

Para 2 :-

sample preparation group 1 : XGBOOST.

i) Define Dataset path in code.

ii) split Data into training & testing sets.

iii) set max iterations = 20

iv) Empty list to initialized to store accuracy values.

v) append value.

Para 3 :-

sample preparation group 2 : LightGBM.

i) Define Dataset path in code.

ii) split Data into training & testing sets.

iii) set max iteration = 20

iv) Empty list is initialized to store accuracy values.

v) append value.

2023/12/20 08:33

Para 4 :-

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

Testing procedure : Run python code in colab.

& Each model trained for 50 Epochs.

Para 5 :-

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

Para 6 :-

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

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.

2023/12/20 08:33

Discussion framework:

Para 1

Result summary:-

while XGBoost delivered higher accuracy, Light GBM demonstrated faster training times and lower computational cost. This advantage becomes more significant with large datasets.

Discussion of Findings:-

Discuss potential trade-off between Accuracy & interpretability 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 architectures for rainfall prediction, potentially outperforming both XGBoost & LightGBM in specific scenarios.

2023/12/20 08:33

Overall consensus:

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

Limitations:

XGBoost provides more insights than LightGBM, ~~buth~~ both models lack the interpretability of traditional statistical methods.
Models trained on specific datasets may not generalize well to other location or weather patterns.

Implications:-

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

Future scope:-

* Investigate hybrid models combining XGBoost & LightGBM for enhanced accuracy & efficiency.
* Develop real-time rainfall forecasting system using LightGBM's fast training capabilities.

Conclusion:-

Choosing optimal algorithm depends on specific data & application needs. This research highlights potential of both XGBoost & LightGBM for rainfall accuracy prediction improvement.

2023/12/20 08:34

T-Test

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
ACCURACY	XGB	20	88.6000	3.08477	.68977
	LGBM	20	75.7500	4.26584	.95387

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
ACCURACY	Equal variances assumed	2.133	.152	10.916	38
	Equal variances not assumed			10.916	34.604

Independent Samples Test

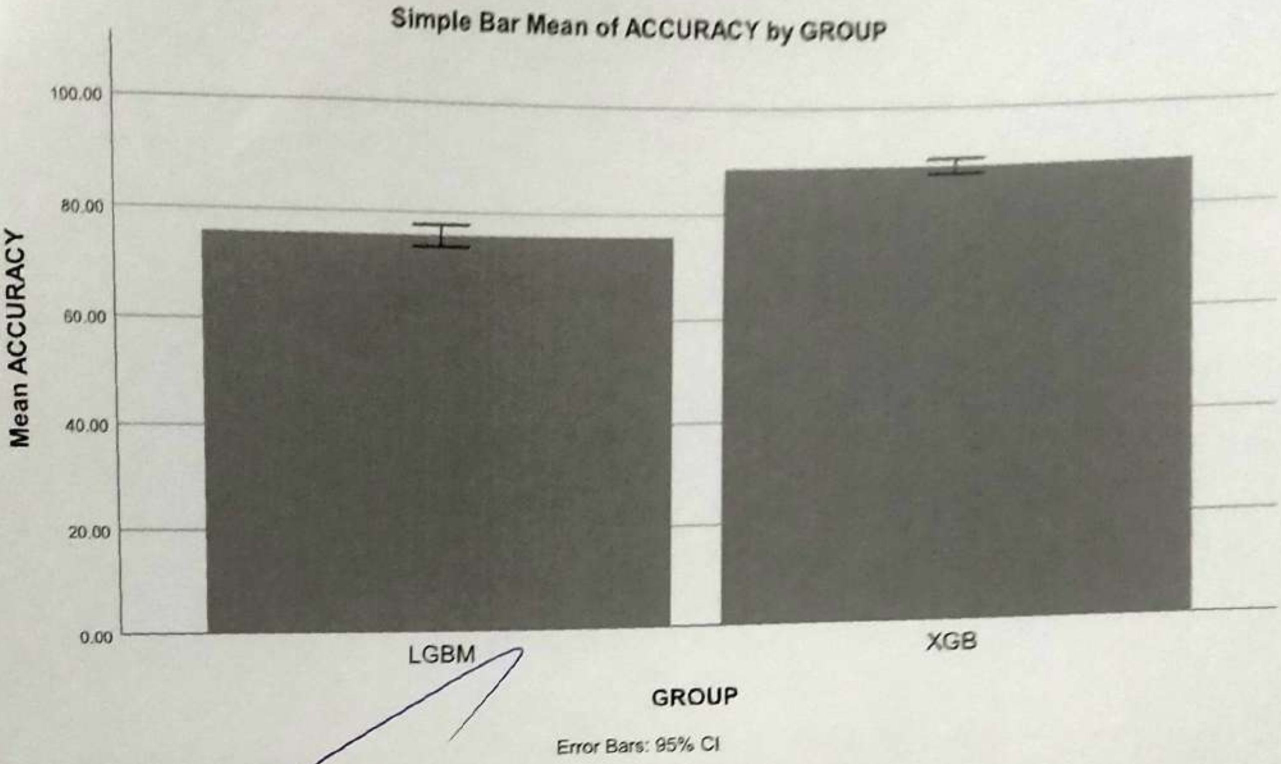
		t-test for Equality of Means		
		Sig. (2-tailed)	Mean Difference	Std. Error Difference
ACCURACY	Equal variances assumed	.000	12.85000	1.17714
	Equal variances not assumed	.000	12.85000	1.17714

Independent Samples Test

		t-test for Equality of Means 95% Confidence Interval of the Difference	
		Lower	Upper
ACCURACY	Equal variances assumed	10.46701	15.23299
	Equal variances not assumed	10.45930	15.24070

GGraph

2023/12/20 08:34



2023/12/20 08:34