Estimation of Rainfall Quantity using Hybrid Ensemble Regression

Preetham Ganesh[1], Harsha Vardhini Vasu[2], and Dayanand Vinod[3]

[1]preetham.ganesh2015@gmail.com, [2]harshavardhini2019@gmail.com, and [3]v_dayanand@cb.amrita.edu

Department of Computer Science and Engineering,

Amrita School of Engineering, Amrita Vishwa Vidyapeetham,

Coimbatore, India

Overview

- Abstract
- Introduction
- Related Works
- Process Flow
- Results and Discussion
- Conclusion

Abstract

- In this paper, ensemble methods are used to predict rainfall in Tamil Nadu, India
- Ensemble Regression Models (ERM) are optimized by tuning various parameters
- On analysis, Bagging Regression model produced better results, but not much in difference
- Ensemble Techniques such as Simple Averaging, Blending and Stacking are used
- Developed models are compared using graphical analysis which is based on actual rainfall values

Introduction

- Rainfall plays a vital role in the every organism on the planet
- It helps in:
  - Maintaining Groundwater table on land
  - Balancing vegetation
- Recently, many cyclones occur at uneven periods, which destroys a lot of things thereby causing more problems to government
- Hence, predicting rainfall will help government and public in many ways

Contribution

- In this paper, we have used ERM to predict rainfall in all regions of Tamil Nadu, India
- The ERM are optimized using empirical analysis which is explained in Results and Discussion section
- Hybrid ERM (H-ERM) are developed by combining various ERM by using ensemble techniques to enhance performance
- Comparison between developed models is done using actual rainfall values
- The best model from previous analysis is compared with similar papers' best model

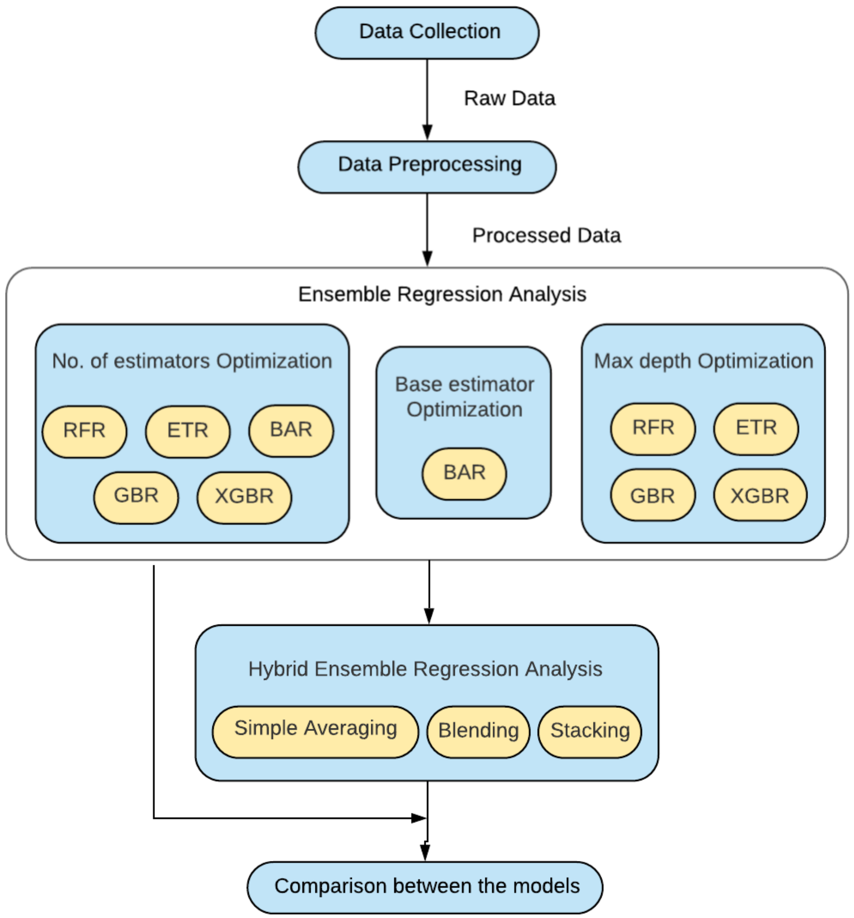
Dataset Description

- The India Water Portal – Met Data Repository was used to collect the data
- Independent Attributes:
  - Average Temperature
  - Cloud Cover
  - Maximum Temperature
  - Minimum Temperature
  - Crop Evapotranspiration
  - Potential Evapotranspiration
  - Vapour Pressure
  - Wet Day Frequency
- Dependent Attribute: Rainfall

Algorithms and Error Measures

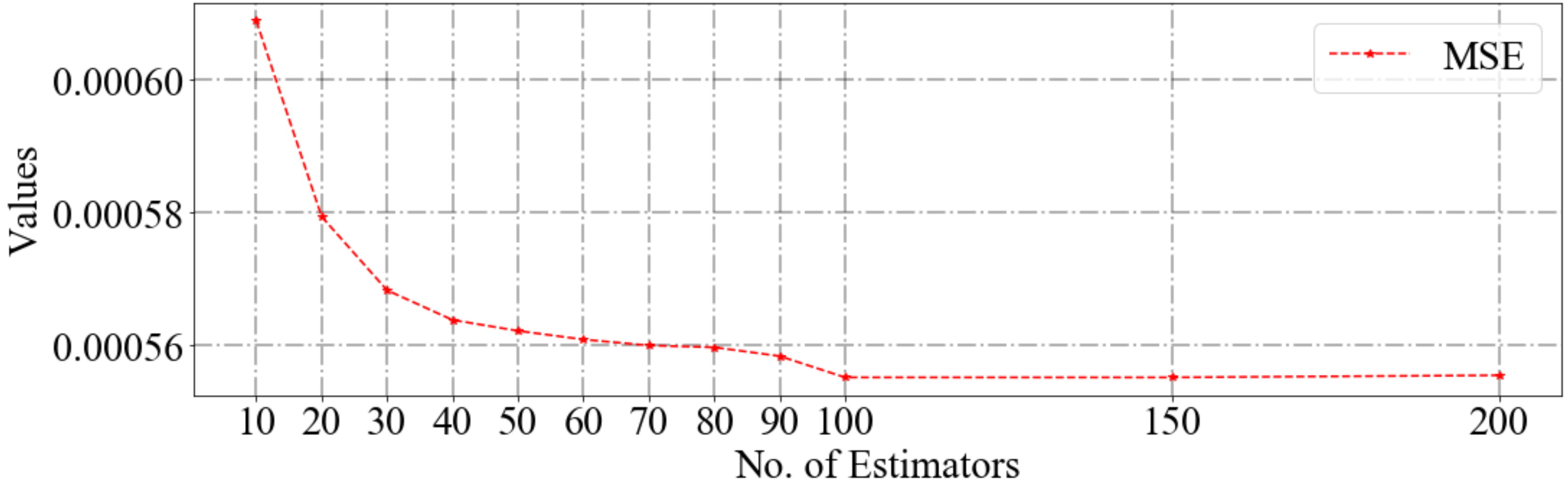
- Regression Algorithms
  - Multiple Linear Regression (MLR)
  - Support Vector Regression (SVR)
  - Decision Tree Regression (DTR)
  - Polynomial Regression (PR)
- Ensemble Techniques
  - Simple Averaging
  - Blending
  - Stacking
- Advanced Ensemble Algorithms
  - Random Forest Regression (RFR)
  - Extra Trees Regression (ETR)
  - Bagging Regression (BAR)
  - Gradient Boosting Regression (GBR)
  - Extreme Gradient Boosting Regression (XGBR)
- Error Measures
  - Mean Squared Error
  - Explained Variance Score

Process Flow



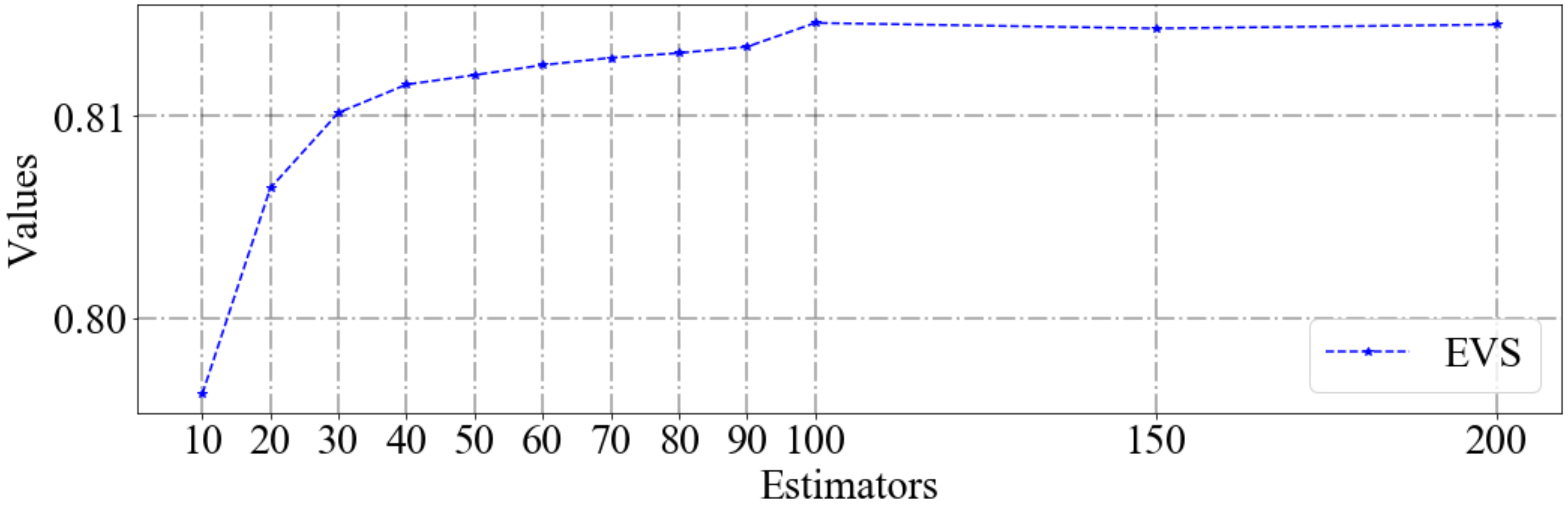
Number of Estimation (NoE) Optimisation

- RFR with different Number of Estimators versus their corresponding MSE values



Number of Estimation Optimisation

- RFR with different Number of Estimators versus their corresponding EVS values



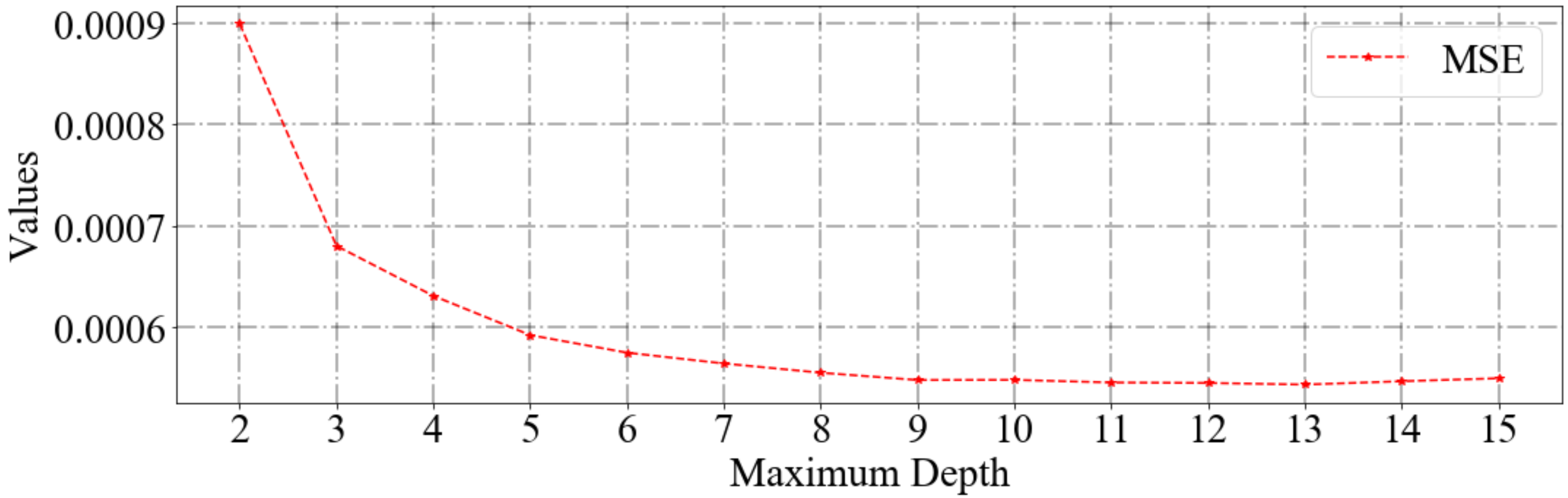
Number of Estimation Optimisation

- Optimal number of estimators for ERM

Ensemble Model	NoE	MSE	EVS
RFR	**100**	**0.000555**	**0.815**
ETR	90	0.000589	0.803
BAR	70	0.000559	0.813
GBR	50	0.000566	0.811
XGBR	50	0.000560	0.813

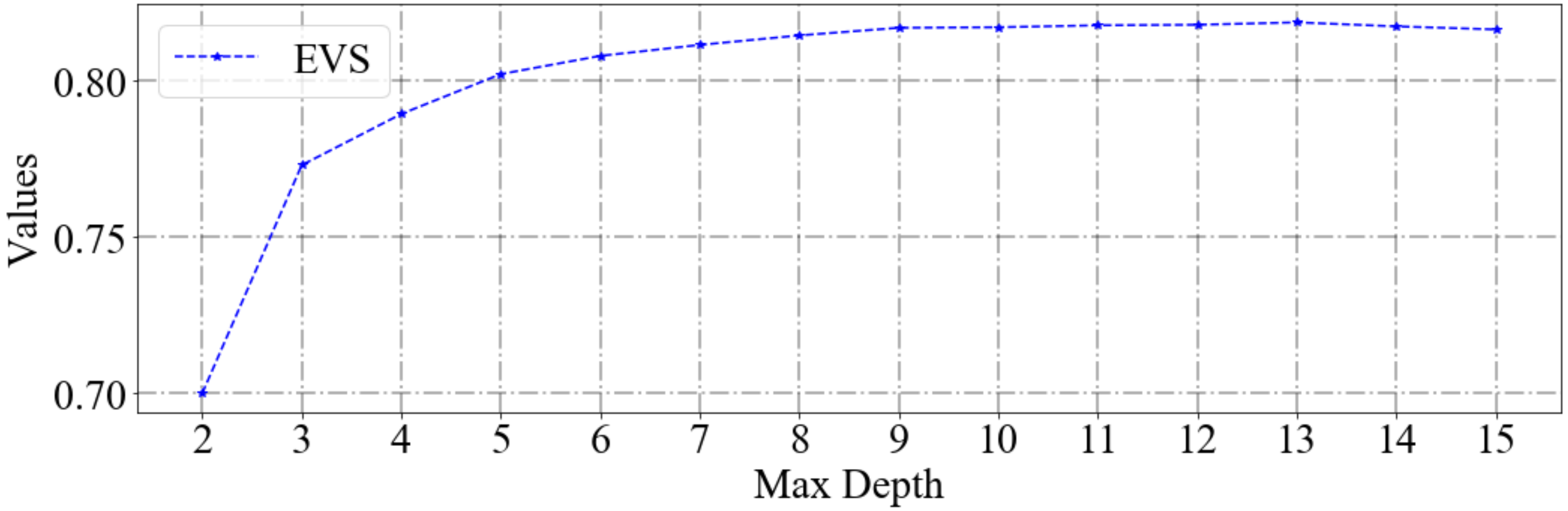
Maximum Depth Optimisation

- RFR with different Maximum Depth versus their corresponding MSE values



Maximum Depth Optimisation

- RFR with different Maximum Depth versus their corresponding EVS values



Maximum Depth Optimisation

- Optimal Maximum Depth for ERM

Ensemble Model	**Maximum Depth**	MSE	**EVS**
RFR	9	0.000548	0.8168
ETR	11	0.000546	0.8178
GBR	**5**	**0.000539**	**0.8199**
XGBR	5	0.000545	0.8190

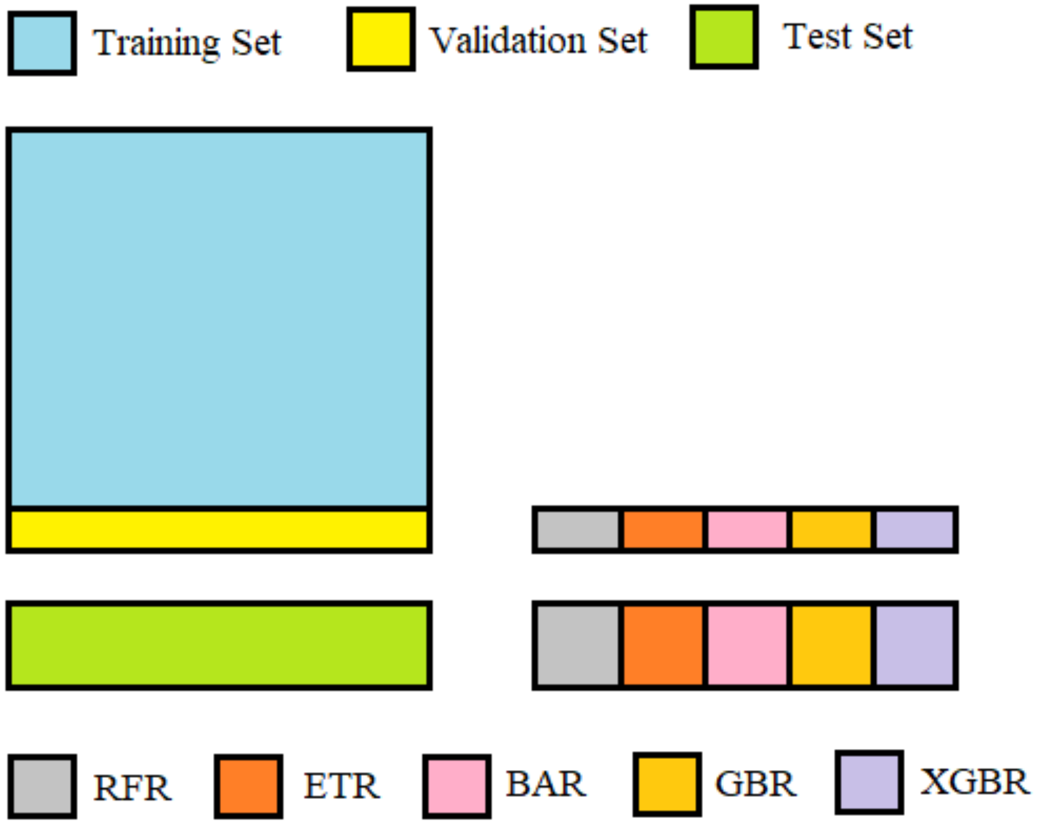
Base Estimator (BE) Optimisation

- Performance of different Base Learners for BAR

Base Learner	MSE	EVS
MLR	0.000646	0.784
DTR[6]	0.000576	0.808
PR[4]	**0.000517**	**0.827**
SVR[L]	0.001447	0.697

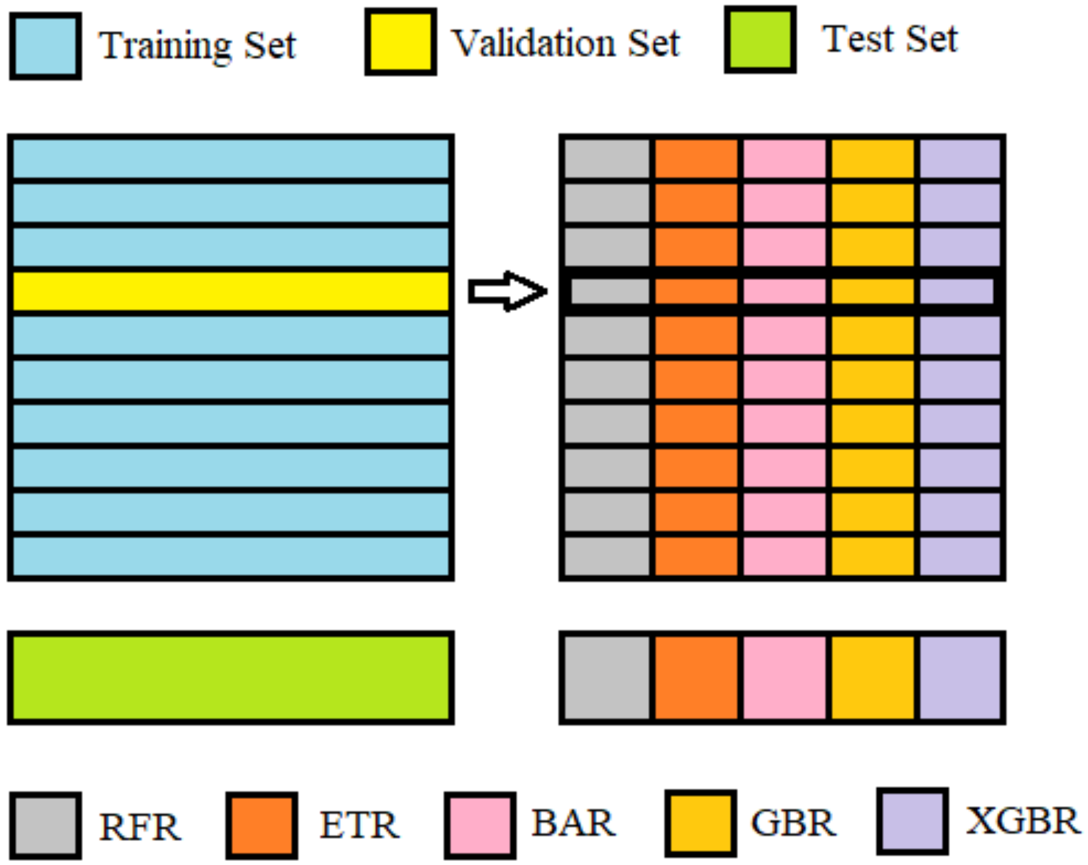
Blending

- Pictorial Representation of Blending based prediction



Stacking (K-Fold Cross Validation)

- Pictorial Representation of Stacking based prediction



Comparison between the H-ERMs using MSE

Combinations	Simple Averaging	Blending				Stacking (Repeat=1)				Stacking (Repeat=10)			
		MLR	PR[4]	DTR[6]	SVR[L]	MLR	PR[4]	DTR[6]	SVR[L]	MLR	PR[4]	DTR[6]	SVR[L]
RFR	0.000544	0.000333	0.000362	0.00048	0.003376	0.000323	0.000314	0.000343	0.0023	0.000323	0.000313	0.000325	0.00209
ETR	0.000556	0.00036	0.000342	0.000463	0.004041	0.000339	0.000323	0.000383	0.002242	0.000337	0.000321	0.000361	0.002385
GBR	0.000539	0.000332	0.000309	0.000456	0.003281	0.000311	0.000298	0.000366	0.002329	0.000312	0.000299	0.000305	0.002213
XGBR	0.000544	0.000319	**0.000298**	0.00044	0.003847	0.000296	**0.000291**	0.000354	0.00232	0.000297	0.000291	0.000301	0.002139
BAR	0.000518	0.00035	0.000499	0.000408	0.002806	0.000344	0.000339	0.000348	0.002	0.000336	0.000332	0.000337	0.001879
RFR, ETR	0.000545	0.000339	0.012075	0.000497	0.003419	0.000322	0.000328	0.000351	0.002042	0.000321	0.000324	0.00033	0.002103
RFR, GBR	0.000536	0.000328	0.003623	0.000483	0.002968	0.000313	0.000316	0.000353	0.002106	0.00031	0.000299	0.00031	0.001999
RFR, XGBR	0.000537	0.000324	0.010304	0.000492	0.003127	0.000301	0.0004	0.00038	0.002219	0.000302	0.000296	0.000316	0.002042
RFR, BAR	0.000515	0.000338	0.010862	0.000426	**0.002011**	0.000306	0.000606	0.000323	0.00183	0.000303	0.00028	0.000316	0.001647
ETR, GBR	0.000538	0.000329	0.000497	0.000485	0.003511	0.000313	0.000308	0.000358	0.002103	0.000309	0.000303	0.000315	0.002219
ETR, XGBR	0.000539	**0.000318**	0.001029	0.000461	0.00343	0.000299	0.000381	0.000338	0.002166	**0.000297**	0.0003	0.000316	0.002242
ETR, BAR	0.000519	0.000357	0.036623	0.000411	0.002397	0.00031	0.000405	0.000326	0.001815	0.00031	**0.000274**	0.000307	0.001846
GBR, XGBR	0.000538	0.000324	0.035758	0.000454	0.003269	0.000296	0.000292	0.000357	0.002217	0.000301	0.00029	**0.000301**	0.002212
GBR, BAR	**0.000513**	0.00036	0.063098	0.000516	0.00268	0.000309	0.000498	0.000319	0.001954	0.000302	0.00028	0.00031	0.001755
XGBR, BAR	0.000515	0.000355	0.004173	0.000431	0.002987	0.000298	0.000835	0.000331	0.001893	0.000297	0.000298	0.000308	0.001763
RFR, ETR, GBR	0.000538	0.000333	0.313515	0.000474	0.00309	0.000313	0.000389	0.000349	0.001945	0.000309	0.000309	0.000309	0.001968
RFR, ETR, XGBR	0.000538	0.000329	0.46725	0.000486	0.003181	0.000301	0.001212	0.000375	0.00206	0.000301	0.000325	0.000329	0.001987
RFR, ETR, BAR	0.000522	0.000341	0.14691	0.000438	0.002098	0.000305	0.001942	0.000315	0.001886	0.000303	0.000288	0.000305	0.001691
RFR, GBR, XGBR	0.000536	0.000323	0.217163	0.000506	0.003105	0.000301	0.000537	0.000381	0.002233	0.000304	0.000291	0.000315	0.001965
RFR, GBR, BAR	0.000518	0.000349	0.061342	0.000431	0.002449	0.000305	0.000531	0.000329	0.001839	0.0003	0.000293	0.000312	0.001733
RFR, XGBR, BAR	0.000518	0.000347	0.136336	0.000444	0.002642	0.000298	0.002017	0.000325	**0.001738**	0.000297	0.000293	0.000309	0.001706
ETR, GBR, XGBR	0.000536	0.000321	1.152639	0.000478	0.003228	0.000297	0.000534	0.000353	0.00223	0.000301	0.000295	0.000315	0.002166
ETR, GBR, BAR	0.000519	0.000361	0.287121	0.000491	0.002552	0.000305	0.000713	**0.000309**	0.001844	0.000301	0.000274	0.000303	0.001798
ETR, XGBR, BAR	0.000520	0.000356	0.430401	**0.000393**	0.002833	0.000297	0.002173	0.000331	0.001864	0.000297	0.000283	0.000304	0.001722
GBR, XGBR, BAR	0.000518	0.000358	0.239132	0.000444	0.002996	0.000297	0.014953	0.000318	0.001877	0.0003	0.000318	0.000306	0.001762
RFR, ETR, GBR, XGBR	0.000536	0.000329	2.621982	0.000525	0.003203	0.000299	0.006256	0.000376	0.002069	0.000304	0.000305	0.000327	0.001947
RFR, ETR, GBR, BAR	0.000522	0.000349	5.742796	0.000431	0.002429	0.000304	0.00074	0.000317	0.001823	0.0003	0.000333	0.000303	0.001692
RFR, ETR, XGBR, BAR	0.000523	0.000346	6.947006	0.000436	0.002672	0.000297	0.006977	0.000314	0.001745	0.000297	0.000291	0.000305	**0.001626**
RFR, GBR, XGBR, BAR	0.000521	0.000348	1.194721	0.000439	0.002715	0.000296	0.013717	0.000331	0.001945	0.0003	0.000342	0.000305	0.001723
ETR, GBR, XGBR, BAR	0.000521	0.000358	18.432151	0.000414	0.002931	0.000295	0.008582	0.00031	0.001857	0.0003	0.000353	0.000303	0.001797
RFR, ETR, GBR, XGBR, BAR	0.000524	0.000347	33.85816	0.00044	0.002756	**0.000295**	0.066363	0.000318	0.002014	0.0003	0.000369	0.000303	0.001692

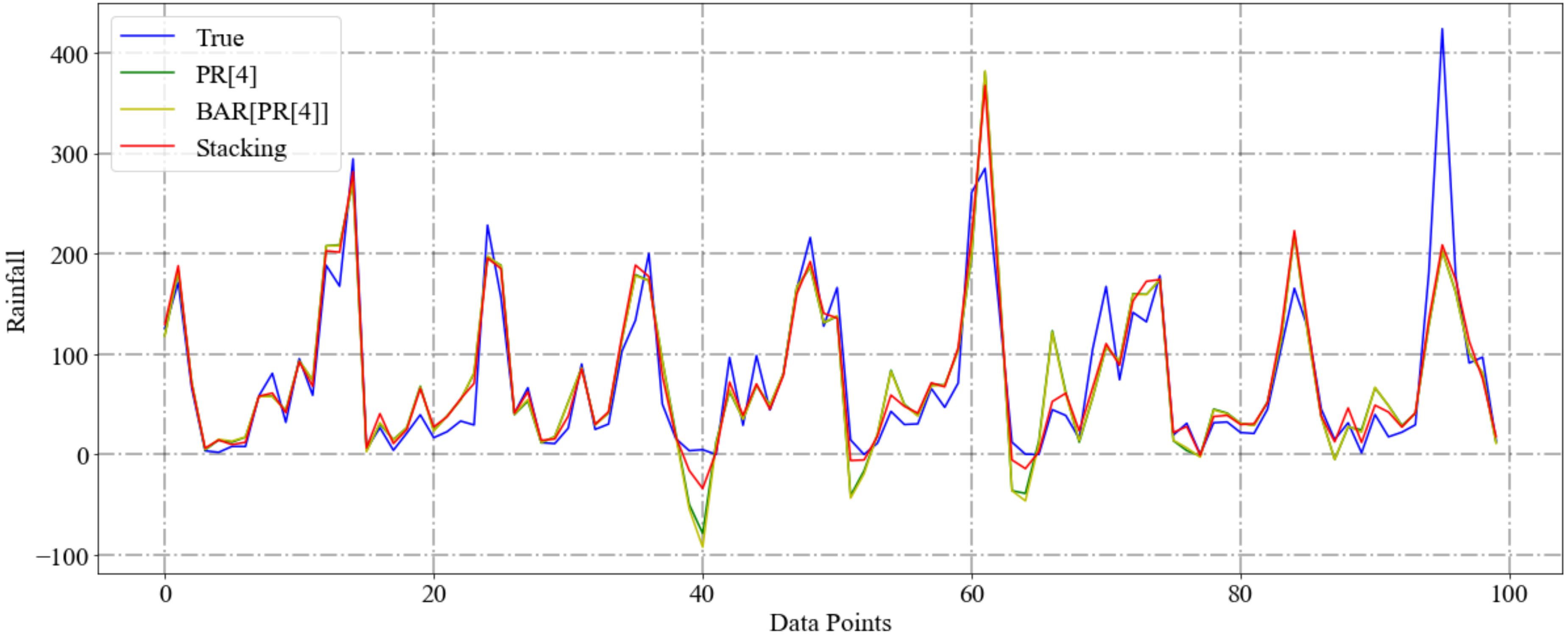
Comparison between the H-ERMs using EVS

Combinations	Simple Averaging	Blending				Stacking (Repeat=1)				Stacking (Repeat=10)			
		MLR	PR[4]	DTR[6]	SVR[L]	MLR	PR[4]	DTR[6]	SVR[L]	MLR	PR[4]	DTR[6]	SVR[L]
RFR	0.8182	0.8924	0.8827	0.8445	0.7257	0.8954	0.8985	0.889	0.8365	0.8956	0.8989	0.8949	0.8432
ETR	0.8143	0.8835	0.8892	0.8501	0.6731	0.8903	0.8956	0.876	0.8332	0.8909	0.8961	0.8831	0.8308
GBR	0.8199	0.8924	0.9	0.8523	0.7258	0.8995	0.9036	0.8817	0.8404	0.899	0.9031	0.9013	0.8397
XGBR	0.8202	0.8968	**0.9036**	0.8575	0.6952	0.9042	**0.9059**	0.8857	0.847	0.9039	0.9058	0.9026	0.8455
BAR	0.827	0.8868	0.8384	0.8681	0.7768	0.8887	0.8902	0.8873	0.8439	0.8912	0.8926	0.8909	0.8488
RFR, ETR	0.8178	0.8904	-2.9083	0.839	0.7475	0.8959	0.894	0.8864	0.8448	0.8963	0.8952	0.8933	0.8428
RFR, GBR	0.8208	0.8939	-0.172	0.8435	0.7659	0.8988	0.8978	0.8858	0.8469	0.8999	0.9033	0.8998	0.8494
RFR, XGBR	0.8209	0.8951	-2.3326	0.8406	0.7463	0.9025	0.8707	0.8771	0.8462	0.9023	0.9043	0.8979	0.8483
RFR, BAR	0.8278	0.8905	-2.5136	0.8622	**0.8212**	0.9008	0.8038	0.8953	0.8578	0.9018	0.9094	0.8978	**0.8642**
ETR, GBR	0.8201	0.8935	0.8391	0.8431	0.7152	0.8988	0.9003	0.8841	0.8483	0.9	0.9021	0.8979	0.8431
ETR, XGBR	0.8202	**0.8972**	0.667	0.8507	0.7218	0.9034	0.8767	0.8907	0.8486	**0.9041**	0.9029	0.8979	0.8453
ETR, BAR	0.8264	0.8845	-10.8559	0.8668	0.8018	0.8997	0.869	0.8945	0.8554	0.8995	**0.9113**	0.9005	0.8533
GBR, XGBR	0.8206	0.8952	-10.5741	0.8529	0.7254	0.9042	0.9058	0.8845	0.8492	0.9026	0.9063	**0.9027**	0.8426
GBR, BAR	**0.8285**	0.8834	-19.4196	0.8331	0.7781	0.9	0.8388	0.8969	0.8525	0.9023	0.9095	0.8997	0.8599
XGBR, BAR	0.8285	0.8851	-0.3508	0.8604	0.7595	0.9035	0.7297	0.8927	0.8598	0.9037	0.9034	0.9003	0.8603
RFR, ETR, GBR	0.8203	0.8922	-100.4767	0.8465	0.7494	0.8988	0.8741	0.8871	0.8517	0.9001	0.9	0.9001	0.8491
RFR, ETR, XGBR	0.8204	0.8936	-150.2365	0.8428	0.7506	0.9028	0.6078	0.8787	0.8497	0.9027	0.8948	0.8937	0.85
RFR, ETR, BAR	0.8255	0.8897	-46.5583	0.8583	0.8129	0.9012	0.3715	0.8982	0.8556	0.9018	0.9068	0.9014	0.8606
RFR, GBR, XGBR	0.8212	0.8954	-69.2848	0.8362	0.7409	0.9028	0.8262	0.8768	0.8457	0.9015	0.9059	0.8981	0.8508
RFR, GBR, BAR	0.827	0.8869	-18.8438	0.8604	0.7929	0.9013	0.828	0.8935	0.8579	0.903	0.905	0.8991	0.8621
RFR, XGBR, BAR	0.827	0.8877	-43.126	0.8564	0.7805	0.9035	0.3474	0.8949	0.8639	0.9037	0.9051	0.9001	0.8636
ETR, GBR, XGBR	0.8266	0.896	-372.123	0.8454	0.7276	0.9038	0.8272	0.8857	0.846	0.9027	0.9046	0.8982	0.8473
ETR, GBR, BAR	0.8266	0.8833	-91.9138	0.8409	0.7826	0.9012	0.7691	**0.9**	0.859	0.9026	0.9113	0.902	0.8552
ETR, XGBR, BAR	0.821	0.8847	-138.3163	0.8726	0.7696	0.9038	0.2966	0.8927	0.8603	0.9038	0.9083	0.9015	0.8593
GBR, XGBR, BAR	0.827	0.8843	-76.3932	0.8563	0.7585	0.9039	-3.8396	0.8972	0.86	0.9029	0.8971	0.9008	0.8587
RFR, ETR, GBR, XGBR	0.821	0.8936	-847.4178	**0.8301**	0.7462	0.9032	-1.024	0.8783	0.8498	0.9018	0.9014	0.894	0.8518
RFR, ETR, GBR, BAR	0.8255	0.8871	-1858.1811	0.8605	0.7931	0.9015	0.7607	0.8973	0.8579	0.903	0.8923	0.9019	0.8612
RFR, ETR, XGBR, BAR	0.8255	0.8878	-2247.591	0.8587	0.7836	0.9038	-1.2588	0.8983	**0.8639**	0.9037	0.9059	0.9013	0.8638
RFR, GBR, XGBR, BAR	0.8261	0.8875	-385.7334	0.8581	0.774	0.9041	-3.4404	0.8929	0.857	0.903	0.8892	0.9012	0.8625
ETR, GBR, XGBR, BAR	0.826	0.8842	-5965.465	0.8659	0.7636	0.9046	-1.7782	0.8996	0.8607	0.903	0.8857	0.902	0.8557
RFR, ETR, GBR, XGBR BAR	0.8251	0.8876	-10960.3019	0.8576	0.7758	**0.9046**	-20.4833	0.8969	0.856	0.903	0.8806	0.9019	0.8612

H-ERMs Analysis

Methods	Performance Measures	Values	Combinations
Simple Averaging	MSE	0.000513	GBR, BAR
	EVS	0.8285	GBR, BAR
Blending [PR[4]]	MSE	0.000298	XGBR
	EVS	0.9036	XGBR
Stacking (Repeat = 1) [PR[4]]	MSE	0.000291	XGBR
	EVS	0.9059	XGBR
Stacking (Repeat = 10) [PR[4]]	MSE	**0.000274**	**ETR, BAR**
	EVS	**0.9113**	**ETR, BAR**

Comparison between H-ERMs using Graphical Analysis



H-ERMs Analysis

Paper Name	MSE	RMSE	EVS
S. K. Mohapatra et al. in [1]	-	9.2433	0.8473
A. H. Manek et al. in [2]	-	0.2060	-
P. Ganesh et al. in [3]	0.00052	0.0227	0.8267
Proposed Model	**0.000274**	-	**0.9113**

Conclusion

- Based on preliminary analysis, it was concluded that BAR with NoE as 70 and BE as Polynomial Regression with degree as 4 performs best
- Various ensemble techniques were used to predict predicted results of ERM, and the developed model was called as H-ERMs
- On analysis it was concluded that, Stacking with repeated K-Fold Cross Validation and base models as ETR and BAR, where:
    - ETR with NoE as 90 and Maximum Depth as 11
    - BAR with NoE as 70 and Base Model as PR with degree as 4

  Performed better than the other models
- The above model performed two times better than all ERMs and H-ERM with Simple Averaging

References

1. Sandeep Kumar Mohapatra, Anamika Upadhyay, and Channabasava Gola. Rainfall prediction based on 100 years of meteorological data. In *2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN)*, pages 162–166. IEEE, Oct 2017.
2. Aishwarya Himanshu Manek and Parikshit Kishor Singh. Comparative study of neural network architectures for rainfall prediction. In *2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, pages 171–174. IEEE, Jul 2016.
3. Preetham Ganesh, Harsha Vardhini Vasu, and Dayanand Vinod. Forecast of Rainfall Quantity and its Variation using Environmental Features. In *2019 Innovations in Power and Advanced Technologies (i-PACT)*. IEEE, 2019, in press.

Questions?

Thank you