Forecast of Rainfall Quantity and its Variation using Environmental Features

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

Department of Computer Science and Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India

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

March 21, 2019

Overview

Abstract

- Rainfall plays a crucial role in the lives of an ordinary man.
- Developing a prediction model that captures sudden fluctuations in rainfall has always been a challenging task.
- The paper aims at developing three models which predict monthly rainfall for all districts in Tamil Nadu, India and also drawing a district-wise comparison among them to find the best model for prediction.
- The models developed are District-Specific Model, Cluster-Based Model and Generic-Regression Model.

Abstract

- The District-Specific Model trains on data from a particular district, the Cluster-Based Model groups districts based on the climatic conditions and trains on data from a particular cluster and the Generic-Regression Model trains on combined data from all the districts.
- The paper also aims at finding the monthly variation of rainfall across geographical regions.

Introduction

- Agriculture is the backbone of India's economy.
- Rainfall is the central source of water for the country's agricultural land.
- It is a boon if the rainfall quantity is in the right amount and a bane if the rainfall is too low or too high where the crops get destroyed.
- The knowledge about the rainfall quantity and its variation can help the farmers to plan their crops, thus saving time, effort and resources.
- These preventive measures can not only save human lives but can also minimise the recovery and reconstruction costs for the state.

Contributions

- The primary focus is on finding the best model among the District-Specific Model, Generic-Regression Model and the Cluster-Based Model along with the best regression algorithm and the corresponding parameter for each district.
- Also, to optimise the result, different parameters for each regression algorithm across all the models are tested.
- Section 6 discusses the variation of rainfall across the geographic regions

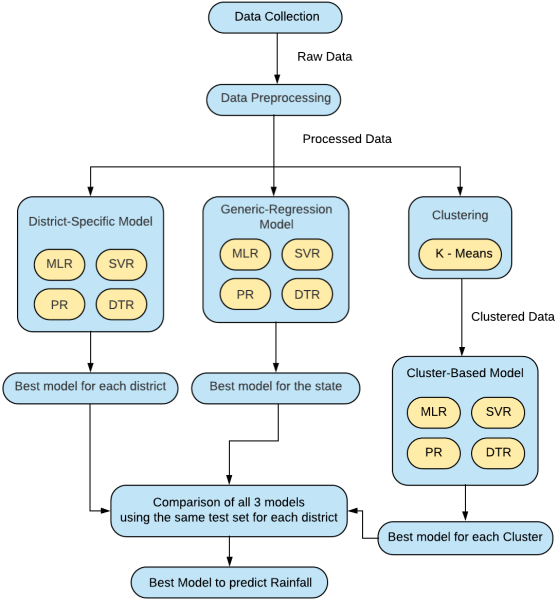
Dataset Description

- The India Water Portal - Met Data Repository is used to collect the data.
- **Independent Attributes**:
  - Average Temperature
  - Cloud Cover
  - Maximum Temperature
  - Minimum Temperature
  - Crop Evapotranspiration
  - Potential Evapotranspiration
  - Vapor Pressure
  - Wet Day Frequency
- **Dependent attribute** - Rainfall

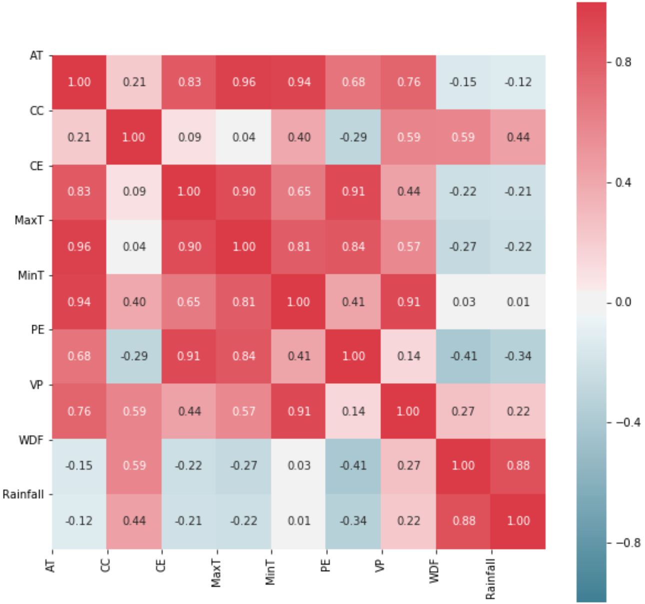
Algorithms and Error Measures

- **Regression Algorithms**
    - Multiple Linear Regression (MLR)
    - Support Vector Regression (SVR)
    - Polynomial Regression (PR)
    - Decision Tree Regression (DTR)
- **Clustering Algorithm**
    - K-Means Clustering
- **Error Measures**
    - Mean Sqaured Error (MSE)
    - Root Mean Squared Error (RMSE)
    - Mean Absolute Error (MAE)
    - Median Absolute Error (MDAE)
    - Explained Variance Score (EVS)
    - R^2 Score (R^2)

Process Flow



Correlation Heat Map



Performance of District-Specific Model

Method	Parameter	MSE	RMSE	MAE	MDAE	EVS	R^2
Multiple Linear Regression	-	0.004	0.0635	0.0442	0.0324	0.8257	0.8241
Polynomial Regression	**Degree = 2**	**0.0039**	**0.0615**	0.0423	0.0282	**0.8305**	**0.8289**
	Degree = 3	0.004	0.0629	**0.0402**	**0.0239**	0.8273	0.8253
	Degree = 4	0.0558	0.1933	0.0846	0.0408	-1.5956	-1.6172
	Degree = 5	4149.7	50.1	15.3	2.7	-187415	-188520
Decision Tree Regression	Max Depth = 2	0.0057	0.0747	0.0484	0.0252	0.7569	0.7552
	Max Depth = 3	0.0043	0.0646	0.0393	0.0198	0.817	0.8156
	Max Depth = 4	0.004	0.0628	0.0371	0.0183	0.8278	0.8264
	Max Depth = 5	**0.0039**	**0.0616**	**0.036**	**0.0181**	**0.8342**	**0.833**
	Max Depth = 6	0.0042	0.0639	0.0368	0.0184	0.8211	0.8199
	Max Depth = 7	0.0044	0.0653	0.0374	0.0184	0.8132	0.812
Support Vector Regression	Kernel = Linear	0.0046	0.0674	0.05	0.0406	0.8053	0.8002
	Kernel = Poly	0.0103	0.1004	0.0727	0.0592	0.5758	0.5637
	Kernel = RBF	**0.0038**	**0.0609**	**0.0424**	**0.031**	**0.8395**	**0.8372**
	Kernel = Sigmoid	0.2638	0.5119	0.3532	0.2414	-10.41	-10.93

Figure: Comparison on performance of the regression algorithms for the Chennai District

Performance of the Generic-Regression Model

Method	Parameter	MSE	RMSE	MAE	MDAE	EVS	R^2
Multiple Linear Regression	-	0.0006	0.0254	0.0156	0.0101	0.7845	0.7844
Polynomial Regression	Degree = 2	0.00057	0.0239	0.0145	0.0089	0.8081	0.8081
	Degree = 3	0.00054	0.0231	0.0137	0.0079	0.8207	0.8206
	Degree = 4	**0.00052**	**0.0227**	**0.0134**	**0.0076**	**0.8268**	**0.8267**
	Degree = 5	0.00053	0.0229	0.0135	0.0078	0.8236	0.8235
Decision Tree Regression	Max Depth = 2	0.00098	0.0313	0.0192	0.0111	0.6731	0.673
	Max Depth = 3	0.00074	0.0272	0.016	0.0091	0.7518	0.7518
	Max Depth = 4	0.00067	0.0259	0.0149	0.0083	0.7759	0.7758
	Max Depth = 5	0.00064	0.0253	0.0145	0.0081	0.7862	0.7862
	Max Depth = 6	**0.00063**	**0.0252**	**0.0142**	**0.0079**	**0.7878**	**0.7877**
	Max Depth = 7	0.00065	0.0254	**0.0142**	**0.0079**	0.7833	0.7833
Support Vector Regression	**Kernel = Linear**	**0.0015**	**0.0388**	**0.0311**	**0.0279**	0.6963	**0.494**
	Kernel = Poly	0.0041	0.0641	0.0574	0.0577	0.5824	-0.3829
	Kernel = RBF	0.0027	0.0523	0.0463	0.0466	0.6845	0.0814
	Kernel = Sigmoid	0.0016	0.0394	0.033	0.0318	**0.7475**	0.4795

Figure: Comparison on performance of the regression algorithms for the Generic-Regression Model

Cluster-Based Model

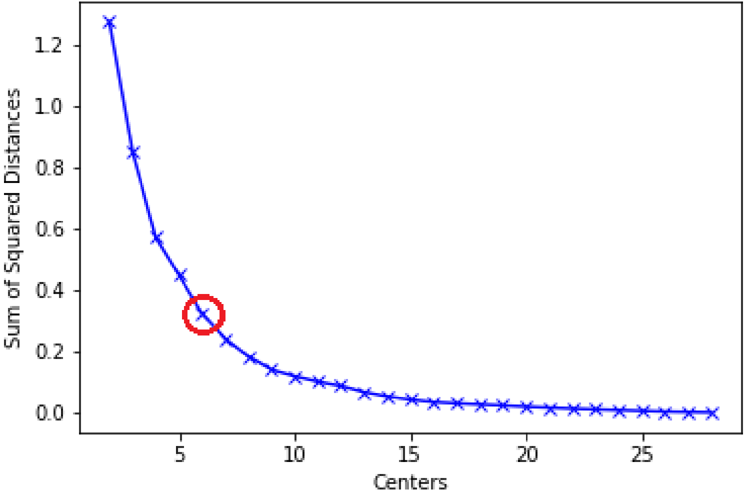
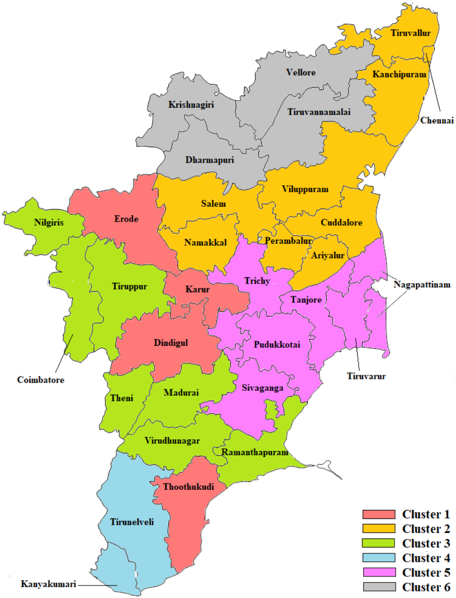


Figure: Elbow Method

Graphical Representation of formed Clusters



Performance of Cluster-Based Model

Method	Parameter	MSE	RMSE	MAE	MDAE	EVS	R^2
Multiple Linear Regression	-	0.0044	0.0663	0.0462	0.0309	0.7412	0.7406
Polynomial Regression	Degree = 2	0.0043	0.0654	0.0453	0.0307	0.748	0.7475
	Degree = 3	**0.0041**	**0.0635**	**0.0437**	**0.029**	**0.7624**	**0.7619**
	Degree = 4	0.0044	0.0659	0.0454	0.0304	0.7444	0.7438
	Degree = 5	0.0106	0.1005	0.0604	0.0.0384	0.379	0.3777
Decision Tree Regression	Max Depth = 2	0.0056	0.0745	0.0534	0.0363	0.6729	0.6722
	Max Depth = 3	0.0048	0.069	0.0478	0.0325	0.7193	0.7187
	Max Depth = 4	0.0047	0.0682	0.0463	0.0309	0.7257	0.7251
	Max Depth = 5	**0.0046**	**0.0678**	**0.0454**	0.0299	**0.7287**	**0.7282**
	Max Depth = 6	0.0048	0.0688	0.0456	**0.0295**	0.7204	0.7198
	Max Depth = 7	0.005	0.0705	0.0463	0.0296	0.7062	0.7056
Support Vector Regression	**Kernel = Linear**	**0.0048**	**0.0694**	**0.0526**	**0.0419**	**0.7343**	**0.7162**
	Kernel = Poly	0.0068	0.0824	0.0675	0.0641	0.6359	0.6001
	Kernel = RBF	0.005	0.071	0.0557	0.0469	0.7292	0.7031
	Kernel = Sigmoid	0.7071	0.8392	0.5115	0.3084	-38.35	-40.84

Figure: Comparison on performance of the regression algorithms for the Cluster 1

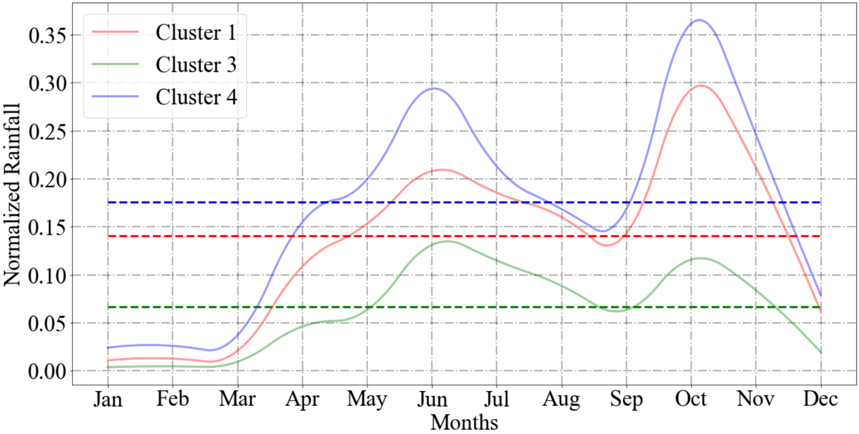
Comparison between the models using MSE, RMSE and MAE

Cluster	District Name	MSE			RMSE			MAE		
		District	Cluster	Generic	District	Cluster	Generic	District	Cluster	Generic
Cluster 1	Dindigul	0.0064	0.0055	**0.0006**	0.0796	0.0734	**0.0245**	0.0559	0.0505	**0.0166**
	Erode	0.0042	0.0026	**0.0003**	0.064	0.0503	**0.0165**	0.0435	0.0338	**0.0109**
	Karur	0.0114	0.0031	**0.0003**	0.1064	0.0555	**0.0184**	0.0778	0.0398	**0.0131**
	Thoothukkudi	0.0074	0.0029	**0.0003**	0.0857	0.0533	**0.0177**	0.0593	0.0369	**0.0121**
Cluster 2	Ariyalur	0.003	0.0008	**0.0001**	0.0539	0.0275	**0.0107**	0.0356	0.0179	**0.0069**
	Chennai	0.0031	0.0022	**0.0003**	0.055	0.0466	**0.0183**	0.0353	0.0287	**0.0112**
	Cuddalore	0.002	0.0005	**0.0001**	0.0441	0.0231	**0.0093**	0.0286	0.0146	**0.0059**
	Kancheepuram	0.0041	0.0017	**0.0003**	0.0634	0.0413	**0.0165**	0.0456	0.0261	**0.0103**
	Namakkal	0.008	0.0012	**0.0002**	0.0891	0.0346	**0.0132**	0.0646	0.0247	**0.0094**
	Perambalur	0.0045	0.0013	**0.0002**	0.0669	0.0352	**0.0134**	0.0453	0.0239	**0.0091**
	Salem	0.0071	0.0011	**0.0002**	0.084	0.0333	**0.0127**	0.0593	0.0229	**0.0087**
	Thiruvallur	0.0046	0.0016	**0.0002**	0.0674	0.0396	**0.0155**	0.0486	0.0253	**0.0098**
	Viluppuram	0.002	0.0005	**0.0001**	0.0441	0.0223	**0.0086**	0.0298	0.0148	**0.0057**
Cluster 3	Coimbatore	0.0035	**0.0009**	**0.0009**	0.059	0.0301	**0.0295**	0.0385	0.0193	**0.0187**
	Madurai	0.007	**0.0008**	**0.0008**	0.0827	**0.028**	**0.028**	0.0546	0.0183	**0.018**
	Ramanathapuram	0.0061	**0.0003**	**0.0003**	0.0776	**0.018**	**0.018**	0.0542	0.0122	**0.012**
	Theni	0.0051	0.0018	**0.0017**	0.0703	0.0412	**0.0411**	0.0433	0.0248	**0.0245**
	The Nilgiris	0.0025	0.002	**0.0019**	0.0486	0.0436	**0.0428**	0.0253	0.0224	**0.0214**
	Virudhunagar	0.0088	**0.0007**	**0.0007**	0.0927	**0.0266**	0.0269	0.0621	0.0176	**0.0175**
Cluster 4	Tirunelveli	0.0082	0.0075	**0.0005**	0.09	0.0856	**0.0228**	0.0611	0.0577	**0.0154**
Cluster 5	Nagapattinam	0.0032	0.0024	**0.0002**	0.0563	0.0489	**0.0151**	0.0378	0.0324	**0.0101**
	Pudukkottai	0.004	0.0019	**0.0002**	0.0625	0.0434	**0.0137**	0.0425	0.0292	**0.0092**
	Sivaganga	0.0048	0.0024	**0.0002**	0.0686	0.0485	**0.0152**	0.0483	0.0337	**0.0105**
	Thanjavur	0.0044	0.0024	**0.0002**	0.0656	0.0483	**0.0149**	0.044	0.032	**0.0098**
	Thiruvarur	0.0054	0.0041	**0.0004**	0.0733	0.0636	**0.0198**	0.0514	0.0426	**0.0132**
	Tiruchirapalli	0.0063	0.0021	**0.0002**	0.0787	0.0458	**0.0143**	0.0576	0.0326	**0.0101**
Cluster 6	Dharmapuri	0.0056	0.0032	**0.0002**	0.0744	0.0566	**0.0125**	0.053	0.0382	**0.0084**
	Tiruvannamalai	0.0054	0.004	**0.0002**	0.0735	0.0628	**0.014**	0.0523	0.0416	**0.0091**
	Vellore	0.0074	0.0049	**0.0002**	0.0854	0.0696	**0.0153**	0.0612	0.0467	**0.0102**

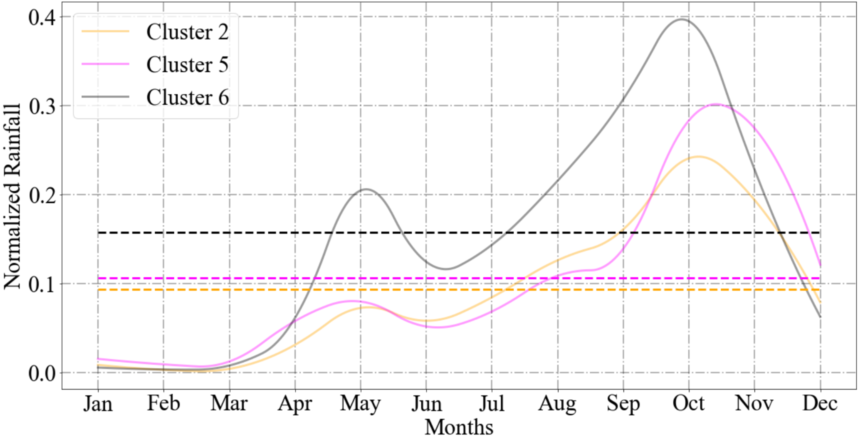
Comparison between the models using MDAE, EVS and R^2

Cluster	District Name	MDAE			EVS			R^2		
		District	Cluster	Generic	District	Cluster	Generic	District	Cluster	Generic
Cluster 1	Dindigul	0.0379	0.0336	**0.0112**	0.7174	**0.7589**	0.7527	0.7146	**0.7571**	0.7508
	Erode	0.0279	0.0212	**0.0069**	0.8407	**0.8654**	**0.8654**	0.8395	0.8643	**0.8646**
	Karur	0.057	0.0268	**0.009**	0.6979	**0.7445**	0.7383	0.6945	**0.7424**	0.7355
	Thoothukkudi	0.0421	0.0251	**0.0081**	0.7018	**0.75**	0.7446	0.6987	**0.7478**	0.7423
Cluster 2	Ariyalur	0.0237	0.0109	**0.0042**	0.8882	**0.9187**	0.9069	0.8873	**0.9181**	0.9062
	Chennai	0.0206	0.0155	**0.0058**	0.8663	**0.9033**	0.8894	0.865	**0.9024**	0.8884
	Cuddalore	0.0186	0.0087	**0.0035**	0.9374	**0.9573**	0.948	0.9369	**0.9569**	0.9475
	Kancheepuram	0.0351	0.0146	**0.0055**	0.8517	**0.916**	0.9016	0.8495	**0.9153**	0.9006
	Namakkal	0.0448	0.0166	**0.0062**	0.7828	**0.8286**	0.8135	0.7803	**0.8273**	0.8117
	Perambalur	0.0283	0.0151	**0.0058**	0.8059	**0.8409**	0.8273	0.8045	**0.8398**	0.8258
	Salem	0.0394	0.0149	**0.0055**	0.7982	**0.8406**	0.8289	0.7963	**0.8391**	0.8274
	Thiruvallur	0.0357	0.0145	**0.0056**	0.8243	**0.9043**	0.8909	0.8227	**0.9035**	0.8901
	Viluppuram	0.0193	0.009	**0.0034**	0.9386	**0.9568**	0.9521	0.9381	**0.9565**	0.9517
Cluster 3	Coimbatore	0.0232	0.0112	**0.0107**	0.8636	0.8847	**0.8888**	0.8625	0.8839	**0.888**
	Madurai	0.0348	0.0115	**0.011**	0.6421	**0.6847**	0.6822	0.639	**0.6814**	0.6797
	Ramanathapuram	0.038	0.0081	**0.0076**	0.7631	**0.8117**	0.8103	0.761	**0.81**	0.8092
	Theni	0.026	0.0145	**0.0138**	0.7353	0.7668	**0.7679**	0.7329	0.7648	**0.7664**
	The Nilgiris	0.0116	0.0099	**0.0085**	0.8494	0.8785	**0.8817**	0.8483	0.8774	**0.8809**
	Virudhunagar	0.0417	0.0115	**0.0112**	0.6459	0.6991	**0.6934**	0.6423	**0.6968**	0.6907
Cluster 4	Tirunelveli	0.0407	0.0388	**0.01**	0.6777	0.7094	**0.7374**	0.6736	0.7069	**0.7348**
Cluster 5	Nagapattinam	0.0247	0.0204	**0.0065**	0.8671	0.8862	**0.8875**	0.8658	0.8853	**0.8866**
	Pudukkottai	0.0275	0.018	**0.0057**	0.8381	**0.8579**	0.8532	0.8368	**0.8567**	0.8523
	Sivaganga	0.0333	0.022	**0.0069**	0.8038	**0.8321**	0.8254	0.8021	**0.8306**	0.8239
	Thanjavur	0.028	0.0198	**0.006**	0.8167	0.8446	**0.8453**	0.8152	0.8434	**0.8442**
	Thiruvarur	0.0359	0.0266	**0.0081**	0.7617	**0.8242**	0.8196	0.7598	**0.8224**	0.8181
	Tiruchirapalli	0.0416	0.0227	**0.0069**	0.7559	**0.8034**	0.8022	0.7528	**0.8019**	0.8
Cluster 6	Dharmapuri	0.0363	0.023	**0.005**	0.7909	**0.8504**	0.8448	0.7894	**0.8493**	0.8439
	Tiruvannamalai	0.0379	0.0251	**0.0053**	0.8366	**0.879**	0.873	0.8352	**0.878**	0.872
	Vellore	0.0448	0.0272	**0.0059**	0.7833	**0.8396**	0.8343	0.7817	**0.8385**	0.8332

Variation of Rainfall across months for Clusters 1, 3 and 4



Variation of Rainfall across months for Clusters 2, 5 and 6



Conclusion

- Based on the analysis, it was observed that the Generic-Regression Model using Polynomial Regression with degree 4 outperforms all the other models and predicts the rainfall in all the districts with comparatively low error rates.
- However, the Cluster-Based Model using Polynomial Regression captures variation in most of the districts and performs better than the Generic-Regression Model only by a fractional value.
- Hence, it can be concluded that Generic-Regression Model is the best model to predict rainfall for the state of Tamil Nadu, India.

References

- WTTC. Country Reports 2017 - India. Technical report, 2017.
- R. Ramanathan and M. Jayakumar. A support vector regression approach to detection in large-MIMO systems. Telecommunication Systems: Modelling, Analysis, Design and Management, 64(4):709717, April 2017.
- J. Isaac and S. Harikumar. Logistic regression within dbms. In 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), pages 661666, Dec 2016.
- Jinghao Niu and Wei Zhang. Comparative analysis of statistical models in rainfall prediction. In 2015 IEEE International Conference on Information and Automation, pages 21872190. IEEE, aug 2015.
- V.P Tharun, Ramya Prakash, and S. Renuga Devi. Prediction of Rainfall Using Data Mining Techniques. In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pages 15071512. IEEE, apr 2018.

References

- Andrew Kusiak, Xiupeng Wei, Anoop Prakash Verma, and Evan Roz. Modeling and Prediction of Rainfall Using Radar Reflectivity Data: A Data-Mining Approach. IEEE Transactions on Geoscience and Remote Sensing, 51(4):23372342, apr 2013.
- Kesheng Lu and Lingzhi Wang. A Novel Nonlinear Combination Model Based on Support Vector Machine for Rainfall Prediction. In 2011 Fourth International Joint Conference on Computational Sciences and Optimization, pages 13431346. IEEE, apr 2011.
- Sandeep Kumar Mohapatra, Anamika Upadhyay. Rainfall prediction based on 100 years of meteorological data. In 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), pages 162166. IEEE, oct 2017.
- Sankhadeep Chatterjee, Bimal Datta, Soumya Sen, Nilanjan Dey, and Narayan C. Debnath. Rainfall prediction using hybrid neural network approach. In 2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications Computing (SigTelCom), pages 67 72. IEEE, jan 2018.

References

- R. Venkata Ramana, B. Krishna, S. R. Kumar, and N. G. Pandey. Monthly Rainfall Prediction Using Wavelet Neural Network Analysis. Water Resources Management, 27(10):36973711, aug 2013.
- Mislan, Haviluddin, Sigit Hardwinarto, Sumaryono, and Marlon Aipassa. Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan - Indonesia. Procedia Computer Science, 59:142151, jan 2015.
- 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 171174. IEEE, jul 2016.
- Yajnaseni Dash, S.K. Mishra, and B.K. Panigrahi. Rainfall prediction of a maritime state (Kerala), India using SLFN and ELM techniques. In 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), pages 17141718. IEEE, jul 2017.