*International Journal of Engineering & Technology*, 7 (4) (2018) 4362-4367

International Journal of Engineering & Technology

*Website: www.sciencepubco.com/index.php/IJET*
*doi: 14125*
***Research paper***

# Analysis of rainfall prediction using machine learning data mining and satellite techniques

**Nikhilkumar B. Shardoor [1] *, Mandapati Venkateswar Rao [2]**

*[1] Research Scholar, Computer Science Engineering, GITAM, Vizag, India*
*[2] Associate Professor, Information Science Engineering, GITAM, Vizag, India*
**Corresponding author E-mail: nikhil053.cse@gmail.com*

**Abstract**

In the era of big data, finding a solution from the huge data is a very big challenge. The buzz word Analytics which helps us in solving such problems, Data analytics is composed of various statistical and analytical methods used to develop new techniques to predict future possibilities. In the current scenario forecasting the rainfall is measured to be an important and thought-provoking task, as it's closely associated with the agriculture, economy and human life. Accuracy of a rainfall forecasting has importance for countries like India whose economy is majorly depends on agriculture. The rainfall prediction is to predict the state of current weather condition. The weather is a dynamic in nature, statistical tech-niques are unsuccessful to provide thee decent accuracy of rainfall. This survey paper provides a different approaches like, Machine Learning, Data Mining and Satellite forecasting's Techniques and their Algorithms are being used and analyzed in getting the better accuracy of rainfall. The comparative study helps experts and non-experts in understanding which method can gives better accuracy of rainfall prediction.

***Keywords***: *Big Data, Classification, Clustering; Data Analytics; Data Mining; Machine Learning; Rainfall Prediction and Regression.*

## 1. Introduction

Data Analytics is a process of Inspecting, transforming and modeling a data (set models and data solutions).The goals of data Analytics are discovering use-full information, suggesting conclusions and supporting decision making. Data analytics is included of solutions used to form analysis models and simulations to create situations, know realities and predict the future. The use of data analytics we can achieve Cost reduction, faster better decision making and new product services. Example: statistics', forecasting, data mining experimental design are used in data analytics.
Rainfall is a random phenomenon its prediction has been always a challenge for the meteorologist all over the world both in terms of complexity and technologically. India's major occupation is agriculture and rainfall is strictly associated with the agricultural region, whereas farming plays a significant role to enrich the state of a country. Effective rainfall is computed considering the different attributes, such as Max Temp, Min Temp, Wind Speed, Humidity, evaporation, Cloud form, Radiation, Sunshine and Rain Fall-Hourly, Monthly, yearly and etc. There are few approaches used to predict the rainfall are, Empirical Method, Physical Method and Statistical Method. Meteorologist uses the traditional method to predict the amount of rainfall by analyzing historical data. (i.e Empirical Method). The accuracy of predicted amount of rainfall is not reached because the main reason for inaccuracy of rainfall are, 1) Rainfall is random event i.e nonlinear and 2) Dependent variable are not constant. (i.e Weather parameters or attributes). Rain fall prediction is useful, 1) Avoid Flood which saves human life and properties. 2) Helps in managing Resource of Water, and 3) Helps Farmer to manage their crops.
In this paper, different algorithms of Machine Learning techniques, Data Mining techniques, and Satellite prediction are analyzed in prediction of rainfall.

## 2. Literature review

In this section, discussed different methods like Machine Learning techniques and Algorithm, Data Mining techniques and Algorithm, and Satellite prediction techniques and how these are all analyzed in prediction of rainfall? What attribute are being used? What is the accuracy of prediction?

### 2.1. Machine learning techniques

Due to Random nature of weather, statistical methods fail to provide good accuracy for Rainfall forecasting. Nonlinearity of data is handled by the Artificial Neural Network techniques and ML Techniques used to forecast better accuracy.

**Table 1:** Different Approaches for Rainfall Prediction Using ML Techniques (2012 To 2014)

| Authors | Region | Dataset Time Period | Techniques | Accuracy Measure | Attributes |
|---|---|---|---|---|---|
| J. Abbot, J. Maroshay (2012) [1] | LOCAL (Australia) | Monthly (1900-2009) | TDRNN | RMSE | Rainfall, Temperature and Solar data. |
| A.Kumar et al. (2012) [2] | LOCAL (Udapi) | Monthly (1960-2010) | EBPNN, CBPNN, LRN | MSE | Average Humidity and Average Wind Speed |
| R. Deshpande (2012) [3] | LOCAL | Monthly | Elman Neural Network | MSE | Rainfall. |
| S. Yeon et al. (2012) [4] | LOCAL | Hourly | Decision Tree, CART | High Prediction Accuracy | Temperature, Wind Direction, Humidity and gust |
| G. Shrivastava et al. (2013) [5] | LOCAL(Ambikapur) | Yearly (1951-2011) | Artificial Neural Network (EBP) | MSE | Humidity, DewPoint, Pressure |
| C. Wu, K Chau (2013) [6] | GLOBAL (India China) | Daily, Monthly | Moving Average, ANN | RMSE | Min MAX Temperature |
| S. Nanda et al. (2013) [7] | GLOBAL (India) | Monthly (1990-2012) | ARIMA Model and ANN | MSE | Min MAX Temperature |
| A.Naik et al. (2014) [8] | GLOBAL (India) | Monthly | Artificial Neural Network (EBP) | RMSE | Temperature, Wind Speed, Humidity |
| Priya et al. (2014) [9] | GLOBAL (India) | Monthly (1871-2010) | Artificial Neural Network (EBP) | RMSE | Min MAX Temperature |
| V.Dhabi, S.Chaudhary (2014) [10] | LOCAL | Daily (1901-2002) | Wavelet ANN | MAE, MSE | Min MAX Temperature, Evaporation, Humidity and Rainfall |
| P.Dutta, H. Tahbilder (2014) [11] | GLOBAL (India | Monthly (6 years) | Regression | RMSE | Min MAX Temperature, Wind Direction Humidity and Rainfall |

## 2.2. Data mining techniques

The data mining approaches uses the two methods, Classification and Clustering Techniques. The following table Table2 represents the DM methods and algorithms applied in rainfall & climate prediction

**Table 2:** DM Methods for Rainfall & Climate Prediction

| Author's | Methods | Algorithms | Data Set Time Period | Accuracy | Attributes | Advantages |
|---|---|---|---|---|---|---|
| S. Yeon et al. (2012) [14] | Decision-Tree. | CART, C4.5 | 3years | 93% | Temp., Wind-Direction, Humidity, gust, Pressure | High Prediction Accuracy |
| S Kannan, S Ghosh [15] | Clustering. Decision-Tree. | K-Mean Clustering. CART. | 50 Years | -- | Wind, rainfall, Temp., Pressure. | Grouping of Multistate rainfall data in clusters |
| F Oliya AB Adeyemo [16] | Decision-Tree, ANN | CART, C4.5. TLFN. | 10 Years | 82% | Evaporation Temp., Wind Speed, Rainfall. | Best Network is predicted for Prediction |
| GJ Sawale [17] | ANN | BPN, Hopfeild Networks | 3 Year | -- | Temperature, Pressure, Wind Speed, Dew point, Humidity. | Combining Both Gives better prediction accuracy. |
| Z jan et al [18] | Lazy Learning | KNN | 10 years | 96% | Dew, Wind-Speed, Humidity, Sea-depth, Rainfall. | Accurate result with huge set of attributes |
| Kalyankar M A, Alaspurkar S. J [19] | Clustering | K- Mean Clustering | 4 Years | -- | Min MAX Temperature, Humidity, Wind Speed, rainfall | Good Prediction Accuracy |
| P.S. Dutta, Tahbilder [20] | Regression | MLR | 6 years | 63% | Min MAX Temperature, Humidity, Wind Speed, rainfall | Acceptable Accuracy |
| M Kannan et al. [15] | Regression | MLR | 5 years | 52% | Min MAX Temperature, Humidity, Wind Speed, rainfall | Can work even with small dataset |

## 2.3. Satellite prediction methods for rainfall

For the growth of precipitation perception various satellite sensor methods or technologies has been developed. The recent growth in precipitation of satellite based techniques are used, which create precipitation items comprising of higher spatial and temporal resolution that is valuable and usable in terms of rainwater assets [21].

**Table 3:** Satellite Prediction Methods for Rainfall

| Techniques | Satellite Data | Features | Future Scope |
|---|---|---|---|
| PERSIANN | TRMM. TMI. GOES8. GOES9/10 GMS5. | Enhance the spatial and temporal determination and exactness of worldwide range precipitation estimation | Stretch out to cover nearly the whole globe in the vicinity of 50°s and 50°w. |
| PERSIANNCSS | GOES. | Enhanced spatial determination. Appraisals rainfall in view of cloud-fix scale. | Evaluate performance over the ocean. |
| CMORPH | GOES8 GOES10 Metosat7 Metosat5 GMS5 | Perform better than PMW precipitation and radar. | Integrate Advanced Microwave Scanning Radiometer for the Earth perceiving system. |

| | | | |
|---|---|---|---|
| | | It is intended to work amid both daytime and evening time conditions. | |
| NRL blended technique | Aqua PMW dataset, GOES9/10/12 Meteosat5/7, Terra. | Enhancing the screening of erroneously recognized small rain over regions of thin cloud.<br><br>Broadens the portrayal of cirrus over a possibility of optical thickness in day and night | Upgrade the utilization of additional multi-spectral strategies and multi-platform watching frameworks for enhancing satellite precipitation valuation.<br><br>Utilization of different kinds of mixed satellite strategy |
| TMPA | LEO<br>DMSP<br>NOAA | Better rich star grouping of satellite allowed precipitation related in both postreal and real-time. | Enhance interalibration of the microwave-based assessments.<br><br>Portray the execution.<br><br>Investigate climatological changes in accordance with the RT items to limit its predispositions |
| CLAVR | NOAA-14 | The exactness of naturally learned choice trees was more prominent than the precision of the cloud covers | NA |
| EPSAT-SG | GEO satellite data | Effortlessly adjusted to other geological territory &operational condition | This strategy can be effortlessly stretched out to territory secured through other G.E.O satellite than MSG2.<br>It can likewise be adjusted to coordinate additional rainfall situation information bases than GPCP |

## 3. Proposed model & algorithm

The proposed predictive model is used for the prediction of rainfall. The predictive model is build using the available rainfall dataset, mathematical equations and algorithms of data mining, machine learning and so on. Very first step is, the dataset is preprocessed for removing unwanted data, noise, and finding the missing values. Once after preprocessing the data, the dataset is divided into two partitions like in training data and testing data i.e 80% of dataset is used for training purpose and 20% of data of dataset is used for the testing the predictive build model. Once after successful validation of the build model i.e the model working efficiently with correct output then the model is deployed for the future application.

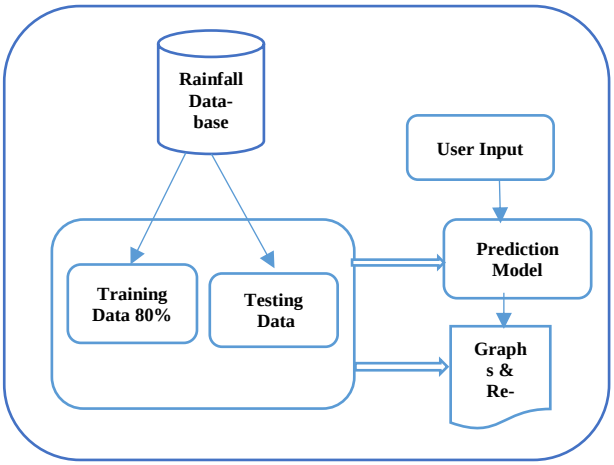


**Fig. 3.1:** Rainfall Forecasting Model.

Algorithm:
The rainfall forecasting techniques or methods are presented as different algorithms. The working of each method and the significant operations are taken out.
Algorithm: Rainfall Prediction
Inputs: Rainfall dataset.
Output: Predicted Rainfall
Start
Step 1: import and accept rainfall dataset.
Step 2: Compute Avg. rainfall.
Step 3: Repeat Steps (3.1 to 3.4) for requested periods
Step3.1: Evaluate rainfall using linear fashion
Step3.2: If (Evaluated rainfall ≥ (Avg. rainfall + 10% Avg. Rainfall)) then
Status ("Above Normal")
Step3.3: If (Evaluated Rainfall ≤ (Avg. rainfall - 10% Avg. Rainfall)) then
Status ("Below Normal")
Step3.4: If (Evaluated rainfall lies between above and below normal limit) then
Status ("Normal")
Step 4: Finally Display the Rainfall status.
Stop

## 4. Results

In this section, descripting the results in the form of graph using the R-Studio. By taking into the consideration of dataset of the annual rainfall in each subdivision of India from 1901 to 2015.
The annual rainfall in each subdivision of India from 1901 to 2015 is shown in following figure. X-> Subdivision and Y-> Annual Rainfall in mm values.

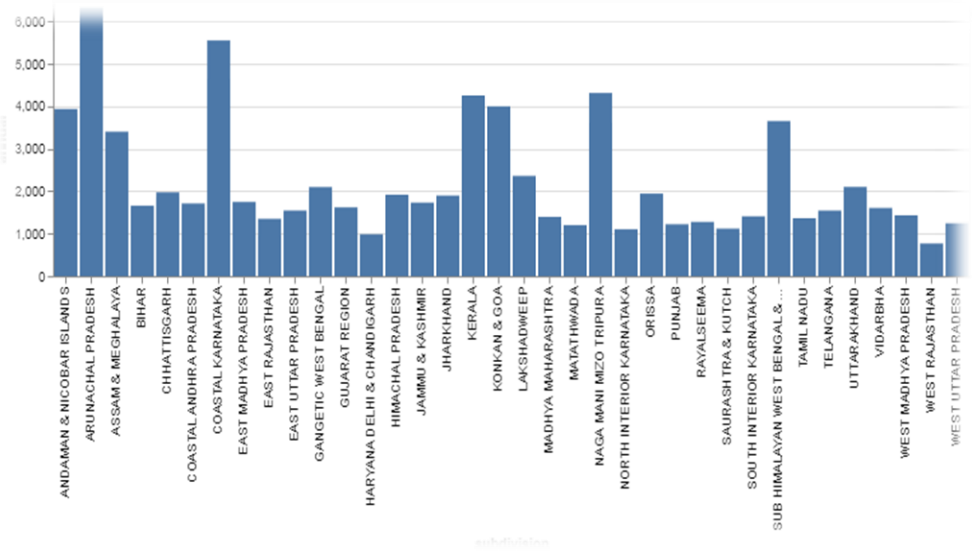


**Fig. 4.1:** Annual Observed Rainfall in Each Subdivision.


**Fig. 4.2:** Analysis Graph of Observed Annual Rainfall.

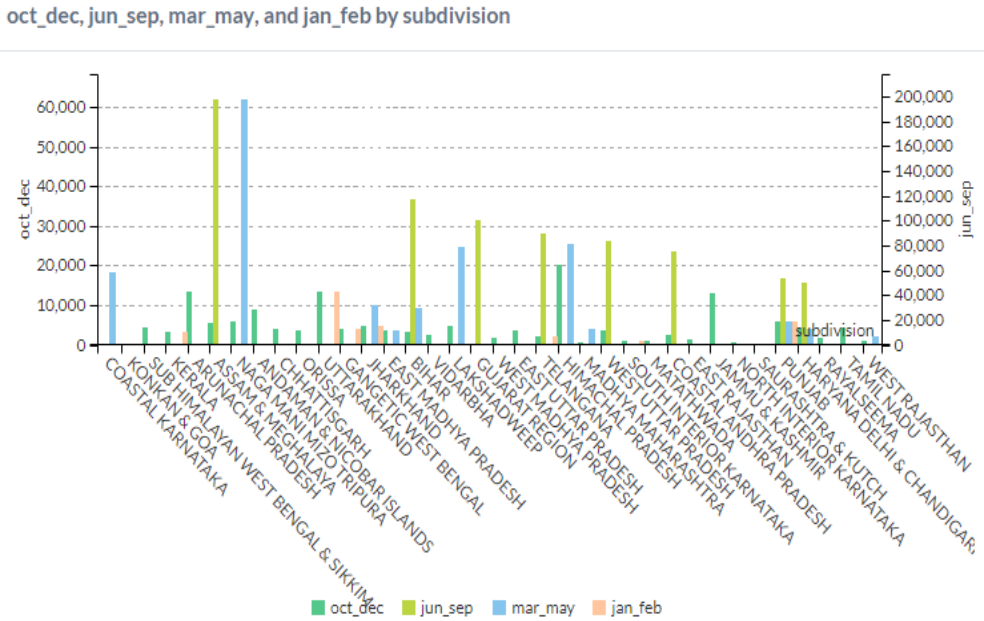


**Fig. 4.3:** Analysis Graph of Rainfall in Jan-Feb, Mar-May, June-Sept and Oct-Dec by Subdivision.

| SUBDIVISION | YEAR | Jan-Feb | Mar-May | Jun-Sep | Oct-Dec |
|---|---|---|---|---|---|
| NORTH INTERIOR KARNATAKA | 1901 | 22.3 | 139.8 | 516.4 | 94.5 |
| NORTH INTERIOR KARNATAKA | 1902 | 0 | 57.1 | 419.3 | 239.6 |
| NORTH INTERIOR KARNATAKA | 1903 | 3.5 | 60.3 | 647.9 | 216.8 |
| NORTH INTERIOR KARNATAKA | 1904 | 0.5 | 65.8 | 425.8 | 86.5 |
| NORTH INTERIOR KARNATAKA | 1905 | 6 | 69.8 | 341.1 | 89.4 |
| NORTH INTERIOR KARNATAKA | 1906 | 21.3 | 32.7 | 560.4 | 169.4 |
| NORTH INTERIOR KARNATAKA | 1907 | 1.1 | 114.6 | 596.7 | 34.3 |
| NORTH INTERIOR KARNATAKA | 1908 | 1.6 | 39.1 | 536.4 | 18 |
| NORTH INTERIOR KARNATAKA | 1909 | 2.5 | 77.1 | 532.5 | 63.6 |
| NORTH INTERIOR KARNATAKA | 1910 | 0 | 35.2 | 587.9 | 136.4 |
| NORTH INTERIOR KARNATAKA | 1911 | 0 | 51.4 | 378.1 | 80.6 |
| NORTH INTERIOR KARNATAKA | 1912 | 12.5 | 80.8 | 532.3 | 133.1 |
| NORTH INTERIOR KARNATAKA | 1913 | 0.9 | 93 | 389.2 | 89.2 |
| NORTH INTERIOR KARNATAKA | 1914 | 0 | 54.2 | 684.5 | 86.1 |
| NORTH INTERIOR KARNATAKA | 1915 | 28.9 | 98.3 | 488.8 | 137.4 |

**Fig. 4.4:** Sample Dataset of Observed Rainfall in North Karnataka.

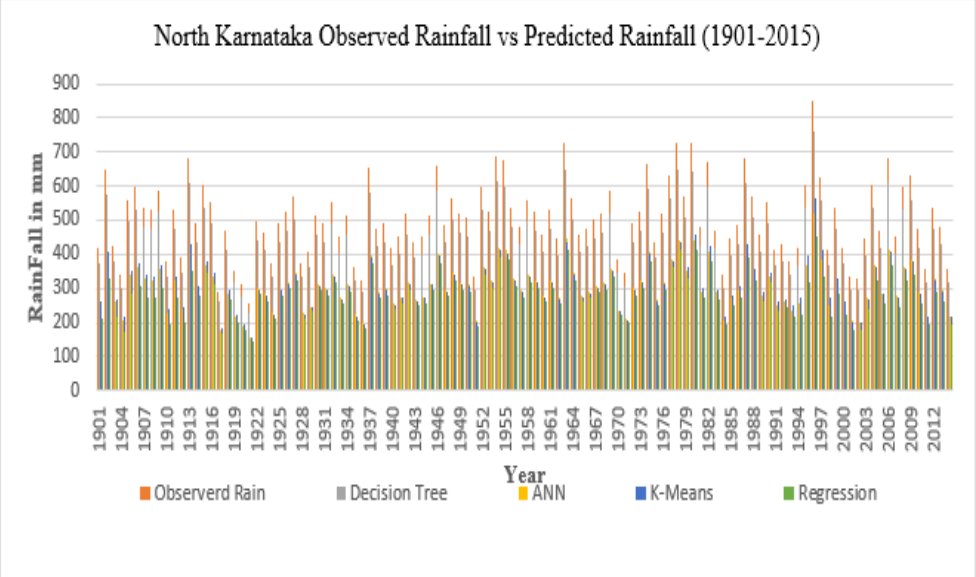


**Fig. 4.5:** North Karnataka Observed Rainfall vs Predicted Rainfall using Decision Tree, ANN, K-Means and Regression.

## 5. Conclusion

The forecasting of rainfall is a very important factor in terms of water resource management, human life and their environment and prior to the agriculture for proper crop management. As rainfall is a nonlinear in nature, it values are not constant, so statistical model yield poor inaccuracy in result. In the process of survey paper investigation of different prevalent Machine Learning, Data Mining and Satellite forecasting techniques and algorithm are presented to predict the rainfall. These techniques would help in predicting the accurate rainfall.

## References

[1] Abbot J and Marohasy J, "Application of ANN to rainfall forecasting in Queensland", Australia, Advances in Atmos. Sci, vol.29, no.4, pp.717-730, 2012. https://doi.org/10.1007/s00376-012-1259-9.

[2] A.Kumar, Ranjan.R and S.Kumar, "A rainfall prediction model using artificial neural network", ICS'GRC, pp. 82-87, 2012.

[3] Deshpande R R, "On the rainfall time series prediction using Multilayer Perceptron Artificial Neural Network," Int. Jr. of Emerging Technology and Advanced Eng., vol.2, no1, pp.148-153, 2012.

[4] Sharad Sharma, SoYeon Ji, Yu Byunggu, Jeong Dong Hyun, "Designing a Rule-Based Hourly Rainfall Prediction Model", IEEE IRI 2012, August – 2012.

[5] G Shrivastava, S Karmakar and M K Kowar, "BPN model for long range forecast of monsoon rainfall over a very small geographical region and its verification for 2012".

[6] K.W.Chau and C.L.Wu, "Prediction of rainfall time series using modular soft computing methods," Engg. Appl. in AI, vol. 26, no. 3, pp. 997-1007, 2013.

[7] Nanda.S.K., Tripathy D P, Nayak S K, and S Mohapatra, "Prediction of rainfall in India using ANN models," Int. Jr. of Intel. Sys. In addition, Appl., vol.5, no.12, pp.1-22, 2013.

[8] S.K.Pathan and A.R.Naik "Indian monsoon rainfall Classification and Prediction using Robust Back Propagation Artificial Neural Network," Int. Jr. of Emerging Technology and Advanced Engg., vol.3, no.11, pp.99-101, 2013.

[9] V.Singh, Priya, Shilpi, Vashistha "Time Series Analysis of Forecasting Indian Rainfall," Int. Jr. of Innovations & Advancement in Comp. Sci., vol.3, no.1, pp.66-69, 2014.

[10] V.K.Dabhi and S.Chaudhary, "Hybrid Wavelet-Postfix-GP model for rainfall prediction of India," Advances in AI.pp1-11, 2014.

[11] D. Pinky Saikia and Tahbilder Hitesh, "Prediction of Rainfall Using Data mining Technique over Assam", I.J.C.S.E, Vol. 5 No.2 Apr May 2014.

[12] Kuo-lin Hsu, B.Imam, S.Sorooshian and Yang Hong, "Global Precipitation Estimation from Satellite Image Using Artificial Neural Networks", Cambridge University Press, pp.21-28.2008.

[13] Folorunsho-olaiya and Adesesan-Barnabas-Adeyemo,"Application of Data Mining Tech. in Weather Prediction and Climate Change Studies", I.J. Inf. Engg and Electronic Business, https://doi.org/10.5815/ijieeb.2012.01.07.

[14] Soo-Yeon Ji, Sharad Sharma, Byunggu Yu and Dong Hyun Jeong, "Designing a Rule-Based Hourly Rainfall Prediction Model, IEEE IRI, August 8-19, 2012.
[15] Pham Huy Thong, Le Hoang Son and Nguyen Dinh Hoa, "Weather casting from satellite image sequences using picture fuzzy clustering and Spatio-Temporal regression", international symposium on geo informatics for spatial infrastructure development for earth and allied sciences, PP: 1-6,2014
[16] F.AI-Roby, Alaa M.EHalees, Data Mining Techniques for Wind Speed Analysis", J. C.E, ISSN: 2010-1619, Vol-2, No: 1, PP:1-5, 2011.
[17] "Decision Support System for Agricultural Management Using Prediction Algorithm", 2013. Mark Ian Animas, Yung-Cheol Byun, Ma.Beth Concepcion and Bobby D. Gerardo,
[18] Charaniya.N.A. and.Dudul.S.V, "Focused Time delay neural network model for Rainfall Prediction Using Indian Ocean Dipole Index", fourth International Conference on Computational Intelligence and Communication Networks, 2012.
[19] Wang Yaming, Wan Dingsheng, Nan Gu and Yufeng Yu, "A Novel Approach to Extreme Rainfall Prediction Based on Data Mining", second Int. Conference on Computer Science and Network Technology, 2012.
[20] T.Mehrnoosh and S. Hashemi, "A Data-Mining Paradigm to Forecast Weather Sensitive Short-Term Energy Consumption", Artificial Intelligence and Signal Processing, Volume: four, Issue No: 4673-1479, PP: 579-584, 2012.
[21] K.L.Hsu, X.Gao, S.Sorooshian, H.V.Gupta, B.Imam, and D.Braithwaite, "Evaluation of P.E.R.S.I.ANN system satellite based estimates of tropical rainfall", Bull.Amer Meteorol.Soc., vol.81, p.2035, 2000. https://doi.org/10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2.
[22] J. L. Berral-Garcia, "A quick view on current techniques and machine learning algorithms for big data analytics", 18th International Conf. on Transparent Optical Networks, pp.1-4, 2016. DOI: 10.1109/ICTON.2016.755051
[23] J. Qui, Q. Wu, G. Ding, Y. Xu and S. Feng, "A survey of machine learning for big data processing", EURASIP Journal on Advances in Signal Processing, Springer, vol. 2016:67, pp. 1-16, 2016. https://doi.org/10.1186/s13634-016-0355-x.
[24] P. Y. Wu, C. W. Cheng, C. D. Kaddi, J. Venugopalan, R. Hoffman and M. D. Wang, "–Omic and Electronic Health Record Big Data Analytics for Precision Medicine", IEEE Transactions on Biomedical Engineering, vol. 64, issue 2, pp. 263-273, 2017. DOI: 10.1109/TBME.2016.257328
[25] M. R. Bendre, R. C. Thool and V. R. Thool, "Big data in precision agriculture: Weather forecasting for future farming", International Conf. on Next Generation Computing Technologies, pp. 744-750, 2015. https://doi.org/10.1109/NGCT.2015.7375220.
[26] Singh, P. and Borah, B., 2013. Indian summer monsoon rainfall prediction using artificial neural network. 27(7), pp.1585-1599
[27] Nikhil Sethi et al, "Exploiting Data Mining Technique for Rainfall Prediction" in International Journal of Computer Science and Information Technologies ISSN:0975 9646 Vol. 5 (3) , pp. 3982-3984, 201
[28] Elia G. P., 2009, "A Decision Tree for Weather Prediction", Universitatea Petrol-Gaze din Ploiesti, Bd. Bucuresti 39, Ploiesti, Catedra de Informatică, Vol. LXI, No. 1
[29] C. L. Wu, K. W. Chau, and C. Fan, "Prediction of rainfall time series using Modular Artificial Neural Networks coupled with datapreprocessing techniques," J. of hydrology, vol. 389, no. 1, pp. 146167, 2010.