Heuristic Prediction of Rainfall Using Machine Learning Techniques

Chandrasegar Thirumalai, IEEE Member
School of Information Technology and Engineering,
VIT University, Vellore, India.
chandru01@gmail.com

M Lakshmi Deepak
School of Information Technology and Engineering,
VIT University, Vellore, India.
lakshmi.deepak2013@vit.ac.in

K Sri Harsha
School of Information Technology and Engineering,
VIT University,
Vellore, India.
sriharsha.2013@vit.ac.in

K Chaitanya Krishna
School of Information Technology and Engineering,
VIT University,
Vellore, India.
1995chaitanyakrishna32@gmail.com

Abstract— **This paper is carried on the heuristic prediction of rainfall using machine learning techniques. As we know agriculture was the predominant of our country and economy. While a regular rain pattern is usually played vital for healthy agriculture but too much rainfall or too little rainfall can be harmful, even it led to devastating of crops. This paper discusses the rate of rainfall in previous years according to various crops seasons like rabi, Kharif, zaid and predicts the rainfall in future seasons. The paper also measures the different categories of data by linear regression method in metrics for effective understanding of agriculture in India. We have selected a real dataset which consists of past year's rainfall rate according to various seasons. Results of this application help farmers to make a correct decision to harvest a particular crop accordingly to crops seasons. Linear regression helps to find**

Keywords: Machine Learning, Rainfall prediction, Regression, Data Mining

I. INTRODUCTION

India's foremost occupation is agriculture and its economy depends upon the agriculture of the country. But rainfall has a dramatic effect on agriculture. Hence, early prediction of rainfall is important for the better growth of the economy. Early prediction of rainfall has been one of the most challenging tasks in the world from past years. Various kinds of techniques used for prediction are Artificial Neural Network (ANN), Support Vector Machine (SVM), Regression analysis, and clustering etc. But the crop seasons in India and two others countries Pakistan and Bangladesh were categorized them into three main seasons as rabi, Kharif, and zaid or zayad. The terms are originated from Arabic where rabi means spring, Kharif means autumn and zaid means summer. Periods of this season are Kharif from July to October during the time of southwest monsoon, a rabbi from October to march and zaid from march to July it's mainly during summer. Each season has its own harvesting crops were in rabi includes wheat, oats, onion, tomato, potato, peas, barley, linseed, mustard oil seeds and masoor etc., Kharif includes rice, sorghum, groundnut, jowar, soya bean, bajra, jute, maize, cotton, hemp, tobacco ragi millet, and arhar etc. and zaid includes sugar cane, cucumber, rapeseed, sunflower, rice, cotton, oilseeds, watermelon, and muskmelon etc. Linear regression is a technique for predicting the various value of a dependent variable from an independent variable when the relationship between the variables can be described by a linear model. It is a method of estimating the conditional expected value, for example, one crop season rabbi given the values of some other seasons like Kharif. The variable of interest rabbi is conventionally called the dependent variable.

II. RELATED WORKS

Dhawal Hirani, Nitin Mishra [1] "A Survey on Rainfall Prediction Techniques". This paper mainly discusses the various machine learning techniques used for early prediction of rainfall and some cognitive approaches. Research's mainly explained about approaches to empirical and dynamical Methods. Pinky Saikia Dutta, Hitesh Tahbilder [2] "Prediction of Rainfall Using Data Mining Technique over Assam". In this paper, they have described data mining technique in forecasting monthly Rainfall and traditional statistical technique Multiple Linear Regression. They included Six years data from 2007 to 2012 this was collected locally from Regional Meteorological Center, Guwahati, Assam, India. They measured the performance of this model in adjusted R-squared. The results of this paper show the prediction model based on multiple linear regression.

V.Brahmananda Rao, K. Hada [3] "An experiment with linear regression in for caring of spring rainfall over south Brazil". It is another related area of research which is based on the aspects of predicting rainfall. They proposed an RMSE approach for predicting rainfall in south Brazil. K.Kaviarasu1, P.Sujith and Mr. G. Ayaappan [4] "Prediction of Rainfall Using Image Processing" They proposed an approach to using digital cloud images for rainfall prediction. In this paper, they used digital cloud images rather than satellite images for better prediction of rainfall and by considering the cost factors and security issues. The status of the cloud is derived by using Cloud Mask Algorithm and type of cloud can be found by

using the K-Means Clustering. They predict type of rainfall cloud by analyzing the color and density of the cloud images and the cloud images are stored as JPEG format

Heather G. Moylan [5] "The Impact of Rainfall Variability on Agriculture Production and Household Welfare in Rural Malawi". The author main aim is using negative rainfall shock index to measure the impact of rainfall. It is simple to combine all required information with the rainfall data since it provided at the plot-level. But other models measure the negative rainfall shock impact on welfare utilizes household-level rainfall data. Many of household's models should have the similar rainfall estimating across plots. Conroy, Skoufias and Vinha [6] Researcher's define the impact of climate variability in rural Mexico by describing weather shocks as rainfall or growing degree day's more than one standard deviation from their corresponding means. They examine the impact on household utilization child height for age, per capita and find that current risk coping mechanism. Pablo and Antonio [7] "The Effect of Rainfall Variation on Agricultural Households: Evidence from Mexico". In this paper, authors determine the impact of rainfall on net income and production in rural Mexico using a two-year dataset. They have constructed a metric on rainfall variation using previous data of various weather stations in Mexico.

II. Prediction Through Linear Regression

This paper represents a review of linear regression technique for the early prediction of rainfall. Linear regression analysis is used for predicting the unknown value of a season from the known value of another season. If rabi and Kharif are two related variables, then linear regression analysis helps us to predict the value of Kharif for a given value of rabbi or vice versa. For example, consider two crop yields rabbi as B and Kharif as A which are related variables. Linear regression analyses can construct the regression line of B on A that would make sense and would be able to determine the dependence of crop yield on rainfall. Then we would be able to predict crop yield given rainfall. This would suggest harvest crops and guaranteed a heavy rainfall if you big crops. In basic straight relapse, we foresee scores on one variable from the scores on a moment variable. The variable we are foreseeing is known as the measured variable and is alluded to as Y. The variable we are constructing our forecasts with respect to is known as the indicator variable and is alluded to as X. At the point when there is stand out indicator variable, the forecasting technique is called straightforward relapse. In basic direct relapse, the subject of this segment, the forecasts of Y when plotted as a component of X shape a straight line.A linear regression model is typically stated in the form.

$$J(\theta) = \frac{1}{2m}\sum_{j=1}^{n}(k_{\theta}a^{j} - b^{j})^{2}$$

Here, m is a number of the sample taken.

k_{θ} variant are 0, 0.5, 1, 1.5, 2; *a, b* are the variables;

TABLE I. Numerical Analysis of Linear Regression.

Yr	X	Y	0*X	0.5*X	1*X	1.5*X	2*X
2006	26	176	31293	26794	22645	18844	15392
2007	41	207	43097	34878	27528	21046	15433
2008	27	199	39930	34633	29713	25169	21003
2009	26	201	40612	35447	30633	26171	22059
2010	27	181	33078	28210	23728	19635	15928
2011	29	171	29241	24481	20145	16230	12738
2012	31	210	44320	37852	31894	26446	21508
2013	20	188	35711	32008	28507	25209	22114
2014	25	191	36557	31886	27533	23500	19787
2015	36	211	44816	37480	30800	24774	19404
2016	28	186	34866	29833	25193	20945	17089
		Σ	188012	159964	134249	110867	89817
		$J(\Theta_1)$	8546	7271	6102	5039	4082

In the same way we have performed linear regression analysis model to all possible subsets of seasons in dataset like [(rabi, Kharif), (rabbi, zaid), (Kharif, rabi), (Kharif,zaid), (zaid, rabbi), (zaid, Kharif)] and they are represented in Table II with corresponding its θ value.

TABLE II. Optimized θ Value.

	RABI	KHARIF	ZAID
RABI	0	4082.633,2	105.71,2
KHARIF	212.081,0	0	134.852,0.5
ZAID	26.429,0.5	539.409,2	0

TABLE III. Calculating Mean and Standard Deviation.

Years	RABI	μ-RABI	(μ-RABI)²
2006	26.41	7.24	52.51
2007	41.68	22.51	506.85
2008	27.45	8.28	68.55
2009	26.5	7.33	53.72
2010	27.83	8.66	75.05
2011	29.06	9.89	97.94
2012	31.93	12.76	162.90
2013	20.13	0.96	0.92
2014	25.26	6.09	37.16
2015	36.2	17.03	290.02
2016	28	8.83	77.96
	210.88		1017.55

Prediction of Rainfall in next Quarter with the details process of this is defined below step-by-step. Finding the Mean and Standard deviation of seasons to predict future rainfall another season

Mean, $\mu = \Sigma (RABI_i) / n$

$=19.17$

Standard Deviation, $\sigma = SQRT(\Sigma(\mu\text{-}RABI_i)^2 / n)$
$= 9.61$

Here, *n* is the number of samples.

TABLE IV. PREDICTION THROUGH MEAN.

μ	RABI	KHARIF	ZAID
RABI	0	4082.63	105.71
KHARIF	212.080	0	134.85
ZAID	26.42	539.40	0

Prediction of rainfall in Kharif and zaid season is done by known value of rabi season. In Table IV the rainfall of Kharif and zaid were predicted as 38.34 cm and 38.34. This was done by derived. For example to get the value of Kharif season rainfall we should find the mean of rabi season and multiply with optimal value's θ Kharif.

$$Predicted\ Rainfall = \mu * (Season) * Optimized\ \theta$$

TABLE V. PREDICTION THROUGH STANDARD DEVIATION.

σ	RABI	KHARIF	ZAID
RABI	0	19.24	19.24
KHARIF	0	0	28.56
ZAID	10.965	43.86	0

Prediction of rainfall in Kharif and zaid season is done by known value of rabi season. In Table V the rainfall of Kharif and zaid were predicted as 19.24 cm and 19.24. This was done by derived. For example to get the value of Kharif season rainfall we should find the SD of rabi season and multiply with optimal value's θ Kharif.

$$Predicted\ Rainfall = \sigma * (Season) * Optimized\ \theta$$

III. CONCLUSION

The main objective of this paper is to help the farmers or agriculture workers such that they can do agriculture more smartly in a much better calculated way. Hereby we will conclude that above results using the linear regression method which suggests the lower correlation between the various crop seasons and data results which we were are predicted are solely done based on the previous year's data and we have kept some good amount of work in order to find the inner relationships between the agriculture parameters

REFERENCES

[1] V. Brahmananda Rao ,K. Hada 1994: An experiment with linear regression in forecasting of spring rainfall over south Brazil

[2] K. Hrona_, P. Filzmoserb and K. Thompsonc 2009 : Linear regression with compositional explanatory variables.

[3] Goutami Bandyopadhyay, 2011: The Prediction of Indian Monsoon Rainfall: A Regression Approach.

[4] Fisher R.A., 1921. On the "probable error" of a coefficient of correlation deduced from a small sample. Metron 1: 3–32.

[5] Elumalai Kannan,2011.:Analysis of trends in India's Agricultural growth.

[6] PS Dutta,2010: Prediction of rainfall using datamining technique over assam.

[7] Hiral chandana ,2013 : Data Across the Agriculture Value Chain .

[8] CT Cheng, 2015 Heuristic Methods for Reservoir Monthly Inflow Forecasting.

[9] Griffith D.A., 2003. Spatial autocorrelation and spatial filtering. Springer, Berlin.

[10] Chandrasegar Thirumalai, Senthilkumar M, "An Assessment Framework of Intuitionistic Fuzzy Network for C2B Decision Making", International Conference on Electronics and Communication Systems (ICECS), IEEE & 978-1-4673-7832-1, Feb. 2016

[11] Rodgers J.L. & Nicewander W.A., 1988. Thirteen ways to look at the correlation coefficient. The American Statistician 42 (1): 59–66.

[12] F. Fioravanti, P. Nesi, "A method and tool for assessing object-oriented projects and metrics management," Journal of Systems and Software, Volume 53, Issue 2, 31 August 2000, Pages 111-136

[13] Galton F., 1875. Statistics by intercomparison. Philosophical Magazine 49: 33–46

[14] Chandrasegar Thirumalai, Viswanathan P, "Diophantine based Asymmetric Cryptomata for Cloud Confidentiality and Blind Signature applications," JISA, Elsevier, 2017.

[15] Galton F., 1877. Typical laws of heredity. Proceedings of the Royal Institution 8: 282–301.

[16] Chandrasegar Thirumalai, Sathish Shanmugam, "Multi-key distribution scheme using Diophantine form for secure IoT communications," IEEE IPACT 2017.

[17] Galton F., 1888. Co-relations and their measurement, chiefly from anthropometric data. Proceedings of the Royal Society of London 45: 135–145.

[18] Chandrasegar Thirumalai, Senthilkumar M, "Spanning Tree approach for Error Detection and Correction," IJPT, Vol. 8, Issue No. 4, Dec-2016, pp. 5009-5020.

[19] Galton F., 1890. Kinship and correlation. North American Review 150: 419–431.

[20] Chandrasegar Thirumalai, Senthilkumar M, "Secured E-Mail System using Base 128 Encoding Scheme," International journal of pharmacy and technology, Vol. 8 Issue 4, Dec. 2016, pp. 21797-21806.

[21] Yule G.U., 1897a. On the significance of Bravais' formulae for regression, in the case of skew correlation. Proceedings of the Royal Society of London Ser. A 60: 477–489

[22] Chandramowliswaran N, Srinivasan.S, and Chandra Segar.T, "A Note on Linear based Set Associative Cache address System" International J. of Computer Science and Engg. (IJCSE) & India, Engineering Journals & 0975-3397, Vol. 4 No. 08 / pp. 1383-1386 / Aug. 2012.

[23] T Chandra Segar, R Vijayaragavan, "Pell's RSA key generation and its security analysis," in Computing, Communications and Networking Technologies (ICCCNT) 2013, pp. 1-5

[24] Chandrasegar Thirumalai, Senthilkumar M, Vaishnavi B, "Physicians Medicament using Linear Public Key Crypto System," in International conference on Electrical, Electronics, and Optimization Techniques, ICEEOT, IEEE & 978-1-4673-9939-5, March 2016.

[25] Chandrasegar Thirumalai, "Physicians Drug encoding system using an Efficient and Secured Linear Public Key Cryptosystem (ESLPKC)," International journal of pharmacy and technology, Vol. 8 Issue 3, Sep. 2016, pp. 16296-16303

[26] E Malathy, Chandra Segar Thirumalai, "Review on non-linear set associative cache design," IJPT, Dec-2016, Vol. 8, Issue No.4, pp. 5320-5330

[27] Hoadley, Katherine A., et al. "Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin." Cell 158.4 (2014): 929-944.

[28] Chandrasegar Thirumalai, Senthilkumar M, Silambarasan R, Carlos Becker Westphall, "Analyzing the strength of Pell's RSA," IJPT, Vol. 8 Issue 4, Dec. 2016, pp. 21869-21874.

[29] Chandramowliswaran N, Srinivasan.S, and Chandra Segar T, "A Novel scheme for Secured Associative Mapping" The International J. of Computer Science and Applications (TIJCSA) & India, TIJCSA Publishers & 2278-1080, Vol. 1, No 5 / pp. 1-7 / July 2012

[30] Chandrasegar Thirumalai, "Review on the memory efficient RSA variants," International Journal of Pharmacy and Technology, Vol. 8 Issue 4, Dec. 2016, pp. 4907-4916.

[31] Vinothini S, Chandra Segar Thirumalai, Vijayaragavan R, Senthil Kumar M, "A Cubic based Set Associative Cache encoded mapping," International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 02 May -2015

[32] Chandrasegar Thirumalai, Himanshu Kar, "Memory Efficient Multi Key (MEMK) generation scheme for secure transportation of sensitive data over Cloud and IoT devices," IEEE IPACT 2017.

[33] M.Senthilkumar, T.Chandrasegar, M.K. Nallakaruppan, S.Prasanna, "A Modified and Efficient Genetic Algorithm to Address a Travelling Salesman Problem," in International Journal of Applied Engineering Research, Vol. 9 No. 10, 2014, pp. 1279-1288

[34] Nallakaruppan, M.K., Senthil Kumar, M., Chandrasegar, T., Suraj, K.A., Magesh, G., "Accident avoidance in railway tracks using Adhoc wireless networks," 2014, IJAER, 9 (21), pp. 9551-9556.

[35] Kahlert, Christoph, et al. "Prognostic impact of a compartment-specific angiogenic marker profile in patients with pancreatic cancer." Oncotarget 5.24 (2014): 12978.

[36] Chandrasegar Thirumalai, Anudeep Duba, Rajasekhar Reddy, "Decision Making System using Machine Learning and Pearson for Heart Attack," IEEE ICECA 2017.

[37] Layton, Robert, and Paul A. Watters. "A methodology for estimating the tangible cost of data breaches." Journal of Information Security and Applications 19.6 (2014): 321-330.

[38] P Viswanathan, P Venkata Krishna, "A Joint FED Watermarking System Using Spatial Fusion for Verifying the Security Issues of Teleradiology," IEEE Journal of Biomedical and Health Informatics, Vol.8, Issue 3, pp. 753-764

[39] Chandrasegar Thirumalai, Vignesh M, Balaji R, "Data analysis using Box and Whisker plot for Lung Cancer," IEEE IPACT, 2017.

[40] Kalaiarassan G, Krishan, Somanadh M, Chandrasegar Thirumalai, Senthilkumar M, "One-Dimension Force Balance System for Hypersonic Vehicle an experimental and Fuzzy Prediction Approach," Elsevier, ICMMM - 2017.

[41] Sonchack, John, Adam J. Aviv, and Jonathan M. Smith. "Cross-domain collaboration for improved IDS rule set selection." Journal of Information Security and Applications 24 (2015): 25-40.

[42] Chandrasegar Thirumalai, Abhiram Reddy P, Jaswanth Kishore Y, "Evaluating Software Metrics of Gaming Applications using Code Counter Tool for C and C++ (CCCC)," IEEE International Conference on Electronics, Communication and Aerospace Technology, 2017.

[43] Amintoosi, Haleh, Salil S. Kanhere, and Mohammad Allahbakhsh. "Trust-based privacy-aware participant selection in social participatory sensing." Journal of Information Security and Applications 20 (2015): 11-25.

[44] Chandrasegar Thirumalai, Aarthi B, Abhinaya V, "Analyzing the Forest Fire using Correlation Methods." IEEE International Conference on Electronics, Communication and Aerospace Technology, 2017.

[45] Chandrasegar Thirumalai, Vaishnavi M, Swapna Anupriya Chandhini, "Analysing the Concrete Compressive Strength using Pearson and Spearman," IEEE International Conference on Electronics, Communication and Aerospace Technology, 2017.

[46] Nguyen, Thi Hanh, Yi Wang, and Renfa Li. "An improved ridge features extraction algorithm for distorted fingerprints matching." Journal of Information Security and Applications 18.4 (2013): 206-214.

[47] P Viswanathan, "Fusion of cryptographic watermarking medical image system with reversible property," in Computer Networks and Intelligent Computing 2011 pp. 533-540.

[48] Chandrasegar Thirumalai, Shridharshan R R, Ranjith Reynold L, "An Assessment of Halstead and COCOMO Model for Effort Estimation," International Conference on Innovations in Power and Advanced Computing Technologies, 2017.

[49] P Viswanathan, P Venkata Krishna, "Fingerprint enhancement and compression method using Morlet wavelet," in Inderscience Publishers 2010, Vol. 3, Issue 4, pp. 261-268

[50] Chandrasegar Thirumalai, Kaitha Sai Sree, Haripriya Gannu, "Analysis of Cost Estimation Function for Facebook Web Click Data," IEEE International Conference on Electronics, Communication and Aerospace Technology, 2017.

[51] Baiardi, Fabrizio, et al. "Automating the assessment of ICT risk." Journal of Information Security and Applications 19.3 (2014): 182-193.

[52] Chandrasegar Thirumalai, Monica S, Vijayalakshmi A, "Heuristics Prediction of Olympic Medals using Machine Learning," IEEE International Conference on Electronics, Communication and Aerospace Technology, 2017.

[53] Sun, Chen, Yang Wang, and Jun Zheng. "Dissecting pattern unlock: The effect of pattern strength meter on pattern selection." Journal of Information Security and Applications 19.4 (2014): 308-320.

[54] Chandrasegar Thirumalai, Kanimozhi R, Vaishnavi B, "Data Analysis using Box Plot on Electricity Consumption," IEEE International Conference on Electronics, Communication and Aerospace Technology, 2017.

[55] P Viswanathan, P Venkata Krishna, "Text fusion watermarking in medical image with semi-reversible for secure transfer and authentication," Advances in Recent Technologies in Communication and Computing, 2009. ARTCom'09. pp. 585-589

[56] CT Cheng, 2015 Heuristic Methods for Reservoir Monthly Inflow Forecasting.