Turkish Journal of Computer and Mathematics Education Vol.12 No. 7 (2021), 1501-1505

Research Article

Rainfall Prediction In North Maharashtra Region Using Support Vector Machine

[1]Husain H. Dawoodi, [2]Manoj P. Patil

[1]School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon (425001), Maharashtra, India
Email: hhdawoodi@nmu.ac.in

[2]School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon (425001), Maharashtra, India
Email: mppatil@nmu.ac.in

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 16 April 2021

ABSTRACT: In this paper a model using Support Vector Machine (SVM) for prediction of rainfall in the three districts of North Maharashtra Region has been proposed. Sixteen input metrological parameters such as temperature, wind speed, wind direction, humidity, visibility etc. were collected from 2009 to 2018 and one output parameter precipitation is used in this study. The model is validated using accuracy, precision, and recall. The results showed that SVM is capable to predict accurate results with 82% accuracy. The SVM model was found to be robust and efficient prediction model considering the varied geographical conditions of these three districts.

Keywords: Machine Learning, Rainfall Prediction, SVM

I. INTRODUCTION

Rainfall is key to hydrological cycle, any alteration in its pattern affects the availability of water resources. The extreme events like droughts, floods occur due to extreme changes in the trends of the rainfall. The production of crop is totally dependent on the amount of moisture in the soil, which in turn is dependent upon the ground water level and the amount of rainfall. The rainfall plays an important role for agriculture in the North Maharashtra Region as this region lacks ample of rivers, lake. Due to the unpredictable nature of the rainfall i.e., occurrence of rainfall in non-monsoon season and non-occurrence of rainfall in monsoon season, farmers who are dependent on the rainfall for their agriculture and have to bear enormous losses to their crops. Hence the research on the occurrences of the rainfall is most significant.

More recently, machine learning (ML) algorithms like Support Vector Machines were examined for forecasting of rainfall in various regions. In the study by [1] prediction of daily rainfall in Hoa Binh province, Vietnam was carried out by comparing various machine learning models like Support Vector Machines (SVM) and Artificial Neural Networks (ANN). The authors have collected and used meteorological parameters like wind speed, solar radiation maximum temperature, relative humidity, and minimum temperature for the prediction. The authors have reported that AI models gave good predictions results and found that SVM is the most robust and efficient method for the prediction of rainfall.

In the study by [2], the authors have proposed a hybrid support vector regression model for monthly rainfall forecasting. The hybrid support vector regression model forecasts one month ahead rainfall at two rain gauge locations of Tabriz and Urmia in northwest Iran. The results of the hybrid model were cross validated with the standalone support vector regression and genetic programming-based models. The proposed model has reported to have higher ability to capture the nonlinear nature of the monthly rainfalls as compared to SVR and genetic models in the semi-arid region of Iran.

In [3] the authors have proposed Support Vector Machine (SVM), Random Forest (RF) and Decision Tree (DT) machine learning models to predict the rainfall at Coonoor in Nilgiris district of Tamil Nadu. The authors have collected and used daily recorded values for wind speed, humidity, temperature, wind direction and cloud speed, from Coonoor weather station. The authors have reported that all the machine learning models have the potential to predict the rainfall in the Coonoor region in Nilgiris district.

In [4] the authors have done a comparative study of various regression models like SVM and ANN for the prediction of wet season. The authors have uses daily rainfall and temperature data from 1995 to 1999 for predicting rainfall runoff and 2001-2003 for the prediction of wet season. The results of the SVM were compared with that of ANN and reported that ANN is computationally intensive and SVM is an efficient alternative for prediction of rainfall. It is further reported that SVM gave better accuracy than ANN.

Average daily and monthly rainfall has been predicted by the [5] for the Fukuoka city in Japan. The authors have done comparative study of data-driven machine learning methods namely ANN, MARS, KNN and

SVM. The proposed method uses hybrid method wherein sub-models are constructed with parameter setting and ranked with the variable selection technique.

The authors in [6] has compared SVM with multi-layer back-propagation Neural Network (MLBPNN). The authors have reported that SVM outperformed MLBPNN and proved to be efficient technique in forecasting time series techniques.

The authors in [7] has modelled a hybrid model of Recurrent artificial neural networks (RNNS) and support vector machines (SVMs), namely RSVR, to forecast rainfall in the typhoon periods from Northern Taiwan. Particle swarm optimization algorithm (CPSO) is used to choose the feature parameters of the SVR model. The results reveals that the proposed model yields better forecasting performance and provides a promising alternative for forecasting rainfall values.

II. METHODOLOGY

A. Study Area:

Maharashtra is one of the largest state in India. Maharashtra came into existence on 1st May 1960. Population wise and area wise Maharashtra ranks second in the country. The state compromises of 36 districts. The zone of North Maharashtra lies in Central India on the north-western corner of the Deccan Plateau, in the valley of the Tapi River. It is bounded to the north by the Satpura Range, to the east by the Berar (Varhad) region, to the south by the Hills of Ajanta, and to the west by the northernmost ranges of the Western Ghats. The region is located at 20°15′30″N to 22°03′00″N latitudes and 73°47′00″E to 76°16′00″E longitudes. North Maharashtra region is geographically very large consisting of three districts viz: Dhule, Nandurbar and Jalgaon.

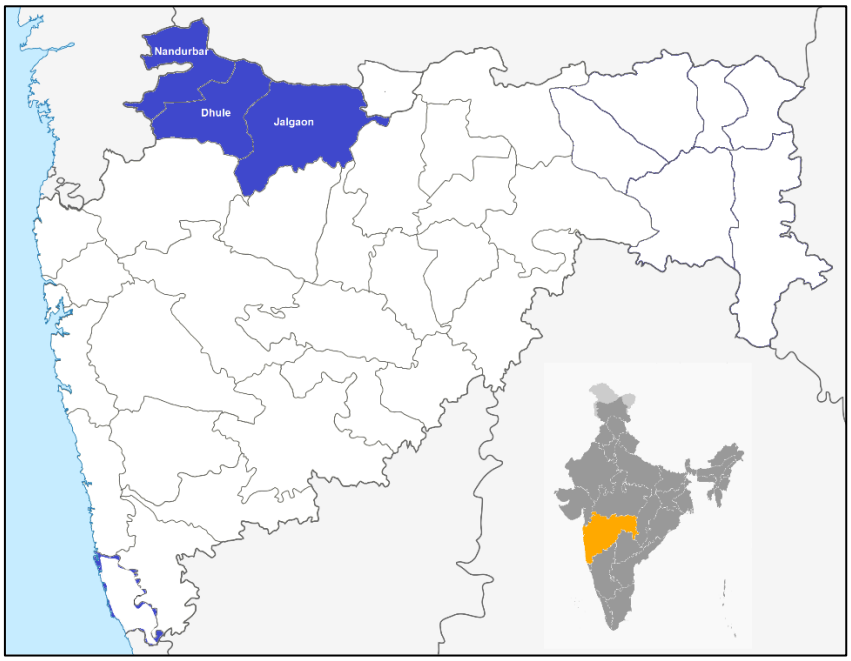

Figure 1: North Maharashtra Region compromising of Jalgaon, Dhule and Nandurbar Districts

B. Data Collection

The hourly metrological data was collected from [8] for the three districts Jalgaon, Dhule and Nandurbar for the period of 2009 to 2018. The list of the predictors is described in table 1.

C. Normalization

The hourly data collected was normalized using Z-Score normalization approach as below:

$$\text{Normalized Data } x' = \frac{x-\mu}{\sigma} \quad (1)$$

Where x is unscaled value and μ is the arithmetic mean and σ is the standard deviation.

$$\text{Arithmetic mean } \mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (2)$$

$$\text{Standard deviation } \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2} \quad (3)$$

Table 1: Description of predictors

Predictors	Lower Bound	Upper Bound	Unit
Moon illumination	0	97	%
Sun hours	7.50	13.70	Hours
Temperature	10	49	°C
Wind speed	0	24	Miles/h
Wind direction	0	360	Degree
Wind directon16Point	--	--	Degree
Precipitation	0.00	49.60	MM
Humidity	8.00	98.00	%
Visibility	2	20	Miles
Pressure	996	1020	Mb
Cloud cover	0	100	%
Heat Index	12	55	°C
Dew Point	-7	30	°C
Wind Chill	12	49	°C
Wind Gust	0	29	Miles/h
Feels Like	12	55	°C

Wind Direction 16 Point Values
N
NNE
NE
ENE
E
ESE
SE
SSE
S
SSW
SW
WSW
W
WNW
NW
NNW

D. Training and Testing

Hourly metrological data for Jalgaon, Dhule and Nandurbar for the period 2009-2018 is considered for analysis purpose. The variable window size approach is used for training the dataset. We have used 2-year to 9-year training window in the training datasets, i.e. past 1-year value is used to predict next year value, or past 8-year values are used to predict the next year value.

E. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm for classification and regression and was developed on the idea of statistical learning theory [9] [10]. SVM finds a hyperplane in an N dimensional space that distinctly separates data points thereby maximizing the distance between two datasets. In SVM data points closer to the hyperplane are called support vectors, which contributes to the determination of the orientation and position of the hyperplane. Input data set must be normalized before the training. This hyperplane is induced by the kernel function K. The distance between the hyperplane and data points is solved by the cost function for solving the problem. Proper selection of Kernel function produces accuracy in least time thereby increasing efficiency of the model. In this research linear kernel function is used.

F. Validation Criteria

The evaluation of the performance of the SVM algorithm is verified by the following indices:

$$\text{Accuracy} = \frac{TP + TN}{N} \tag{4}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

Table 2: Confusion Matrix

	Prediction	
	Precipitation	Non-Precipitation
Observation		
Precipitation	TP	FN
Non-Precipitation	FP	TN

Where the number of events True Positive (TP), False Negative (FN), False Positive (FP), True Negative (TN) are defined by the confusion matrix in Table 2, and N is defined as TP + FN + FP + TN. TN are the number of events where prediction is precipitation when the observation is precipitation. FN are the number of events where prediction is non-precipitation when the observation is precipitation. FP are the number of events where the prediction is precipitation when observation is non-precipitation. TN are the number of events where prediction is non-precipitation when observation is non-precipitation.

III. RESULTS AND DISCUSSION

The extensive experiments were carried out on Intel Xeon E5-2667 @ 3.20 GHz processor with 64 cores having 64 GB of RAM to evaluate the performance of the SVM algorithm for prediction of rainfall in three districts. For predicting the precipitation or no-precipitation the data from the previous years were used for training purpose and the prediction of the next year is done. For predicting the precipitation for 2011, the data from 2009 and 2010 is used for training. Similarly, for predicting the value for 2018 the values from 2009 to 2017 are used for training datasets.

The experiment results for the three districts viz: Jalgaon, Dhule, Nandurbar are shown in Table 3. Overall prediction result for all the three districts is shown in Fig. 2. It can be observed that the SVM model is working well in forecasting rainfall for three different districts of the North Maharashtra Region, these three districts have different geographical conditions ranging from plains and plateaus to hills. It is further observed that as the number of records for training is increased, the accuracy of the model increases. From the results reported in Fig. 2, it is observed that there is a drop in precision level and accuracy of the model. This drop is due to low precipitation in that year resulting in large number of true negative records for that particular year.

The experiments states that SVM is capable to predict accurate results with 82% accuracy. It further proves the ability of application of machine learning algorithm to the predication of precipitation.

Table 3: Prediction results of SVM for Jalgaon, Dhule and Nandurbar Regions

Year	Total Records	Accuracy	Precision
2009	1,85,976	0.754	0.434
2010	3,69,936	0.759	0.503
2011	5,53,896	0.754	0.517
2012	7,38,360	0.771	0.517
2013	9,22,320	0.76	0.503
2014	11,06,280	0.777	0.509
2015	12,88,776	0.769	0.499
2016	14,73,240	0.765	0.429
2017	16,52,785	0.804	0.526
2018	17,75,258	0.826	0.599

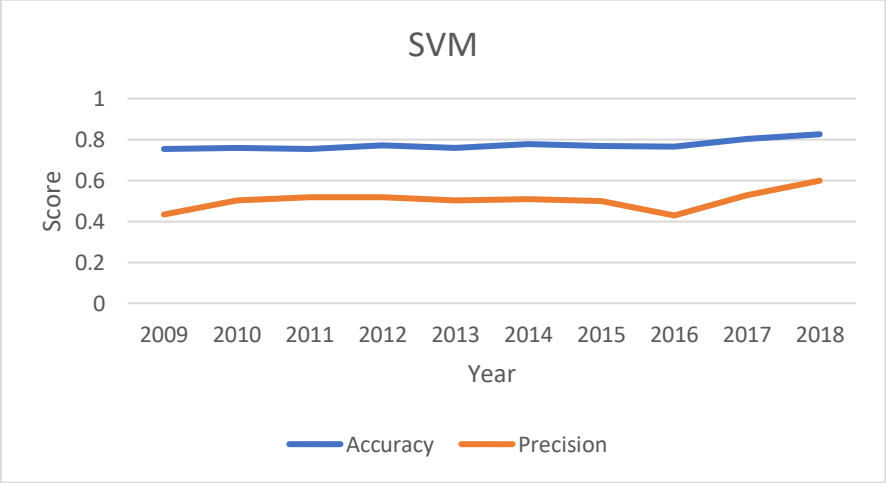

Figure 2: Accuracy and Precision of SVM

IV. CONCLUSION

In the present study Support Vector Machine (SVM) is applied for the prediction of rainfall in the three districts of North Maharashtra Region. Sixteen parameters like temperature, humidity, wind speed, wind direction, pressure etc. were used as input parameters and precipitation was used as output parameter. Validation of the model was done using Accuracy, Precision and Recall. Performance of the proposed model in prediction of rainfall is giving accuracy of 82%. The single SVM model is able to predict rainfall for Jalgaon, Dhule and Nandurbar districts having different geographical conditions.

REFERENCES

1. B. T. Pham, L. M. Le, T.-T. Le, K.-T. T. Bui, V. M. Le, H.-B. Ly and I. Prakash, "Development of advanced artificial intelligence models for daily rainfall prediction," *Atmospheric Research,* vol. 237, p. 104845, 2020.
2. A. Danandeh Mehr, V. Nourani, V. Karimi Khosrowshahi and M. A. Ghorbani, "A hybrid support vector regression--firefly model for monthly rainfall forecasting," *International Journal of Environmental Science and Technology,* vol. 16, pp. 335-346, 01 1 2019.
3. V. P. Tharun, R. Prakash and S. R. Devi, "Prediction of Rainfall Using Data Mining Techniques," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2018.
4. N. Sharma, M. Zakaullah, H. Tiwari and D. Kumar, "Runoff and sediment yield modeling using ANN and support vector machines: a case study from Nepal watershed," *Modeling Earth Systems and Environment,* vol. 1, p. 1–8, 2015.
5. S. M. Sumi, M. F. Zaman and H. Hirose, "A rainfall forecasting method using machine learning models and its application to the Fukuoka city case," *International Journal of Applied Mathematics and Computer Science,* vol. 22, pp. 841-854, 2012.
6. R. Samsudin, A. Shabri and P. Saad, "A comparison of time series forecasting using support vector machine and artificial neural network model," *Journal of applied sciences,* vol. 10, p. 950–958, 2010.
7. W.-C. Hong, "Rainfall forecasting by technological machine learning models," *Applied Mathematics and Computation,* vol. 200, pp. 41-57, 2008.
8. "World Weather Online," [Online]. Available: https://www.worldweatheronline.com/. [Accessed March 2020].
9. C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, p. 273–297, 1995.
10. V. Vapnik, "The support vector method of function estimation," in *Nonlinear modeling,* Springer, 1998, p. 55–85.