International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 02 | Feb 2019 www.irjet.net p-ISSN: 2395-0072

PREDICTION OF WEATHER AND RAINFALL FORECASTING USING CLASSIFICATION TECHNIQUES

N. DIVYA PRABHA[1], P. RADHA[2]

[1] Research Scholar, Department of Computer Science, Vellalar College for Women, Erode
[2] Assistant Professor, Department of Computer Science, Vellalar College for Women, Erode, Tamilnadu, India

---

Abstract - *The extensive rainfall data series is acting as an most important position in all water related studies. Regularity and connection of rainfall data series are incredibly significant for obtaining some valuable or reliable results from such studies. Though, these rainfall data series frequently hold gaps or misplaced values due to different reasons like as the lack of observers, struggle with measuring devices, loss of information or records etc. The utilize of a rainfall data series with lost values may significantly authority the statistical power and correctness of a study. By estimating and extensive the nowhere to be found rainfall data, a series could be through longer to build the water related study additional dependable. Improved Multilayer Perception Neural Network is planned an intellectual tool for predicting Rainfall Time Series. This Rainfall Data series has been approved using the projected Multilayer Perceptron Neural Network. It seems that the presentation process such as MSE (Mean square error), and NMSE (Normalized mean square error) on testing as well as preparation of data set for small term forecast are found as best possible in assessment with other network like as Adanaive, AdaSVM.*

***Key Words*: Classification, svm, Naive Bayes, j48, Artificial Neural Network.**

1. INTRODUCTION

1.1 DATA MINING

Data mining (sometimes called data or knowledge discovery) is the progression of analyzing data from special perspectives and abbreviation into useful data information that can be used to enlarge the revenue, reduce cost and both.

Data mining software is individual and has number of logical tools for analyzing the data. It allows the users to analyze data from various dimensions or angles and review the associations recognized. Technically, the data mining is the process of decision correlations or patterns between fields in huge relational databases. It is the computerized application of particular techniques/algorithms in instructs to recognize particular prototype from the huge data sets. The research work is developed via learning conception by using computer part of area data mining called as machine learning.

1.2 RAINFALL PREDICTION

In bygone days, most of organization did manually process of extraction patterns form data sets. Now a day's modern computer world, collection of data set, planning and its storage is significantly enlarged. India's most major predominant occupation is agriculture, accounting for about 52% of employment. In 1993-94 and 2009 total workforce in service region makes up an additional 34% and industrial region approximately 14% and agricultural accounted 64%. The Agriculture needs mostly sub-continent depend on the rainfall model. Data mining techniques which can be constructive for prediction and decisions more than crop planting in the areas with the help of rainfall data model.

India's eastern part occupies the agricultural areas regularly face with rigorous deficiency every once a three year phase. It is moreover established that a bunch of industries have been occupied place and created high pollution more than the district.

Weather forecasting considered two methods,

(a) The empirical approach

(b) The dynamical approach.

The empirical approach is the first approach and it is based on incidence of analogs and is over and over again referred by meteorologists as analog forecasting. It is much useful approach for predicting local-scale climate conditions if recorded data's are abundant.

The other hand on dynamical approach is based on equations and onward simulations of the impression and is time and again referred to as computer modeling. It is only useful for modeling large-scale weather conditions phenomenon and may not predicts temporary weather proficiently. The majority meteorological processes frequently demonstrate temporal and spatial changeability. To experience of issues of nonlinearity of physical processes, incompatible spatial and temporal range and insecurity in restriction estimates.

2. LITERATURE REVIEW

Cheng Zhou, Boris Cule, Bart Goethals et. al,[2016], The use of patterns in predictive models is a

subject that have been inward a lot of awareness in recent years. Pattern mining can assist to obtain models for prepared domains, such as graphs and sequences, and has been projected as a means to find more perfect and more interpretable models. In the face of the huge amount of publications committed to this area, it consider conversely that a summary of what has been consummate in this area is not there. This work presents our viewpoint on this developing area. The main beliefs of pattern mining that are significant when mining patterns for models and make available an overview of pattern-based classification methods. In catalogue these methods beside the sub sequent proportions: (1) whether they post-process a pre-computed locate of patterns or iteratively implement pattern mining algorithms; (2) whether they decide on patternsmodel-independently or whether the pattern selection is guide by a model. Review the results that have been obtained for each one of these methods.

P.Samuel Quinan, Miriah Meyer et al [2016],Meteorologists evolution and analyze climate forecasts using hallucination in command to observe the behaviours of and associations with climate features. In this intend study conducted with meteorologists in result carry roles, we recognized and attempted to deal with two significant frequent challenges in weather apparition: the employment of conflicting and repeatedly unsuccessful visual encoding practices cross ways a large range of visualizations, and a lack of hold up for straight visualizing how diverse weather description narrate across an collection of potential forecast outcomes. In this effort, present a classification of the exertion and data connected with meteorological forecasting, we intend a set of conversant default programming choices that incorporate existing meteorological conventions with efficient visualization preparation, and we make longer a set of techniques as a primary step toward honestly visualizing the communications of numerous features over an ensemble forecast. We converse the incorporation of these charity keen on a purposeful prototype tool, and as well as imitate on the numerous sensible challenges that occur when working with weather data.

C. R. Rivero, J. Pucheta, S. Laboret, M. Herrera and V. Sauchelli et al [2013],In this occupation an algorithm to fiddle with parameters using a Bayesian method for cumulative rainfall time series forecasting implementedby an ANN-filter is obtainable. The principle of modification comprises to produce a posterior probability distribution of time series standards from forecasted time series, where the organization is transformed by consider a Bayesian deduction. These are approximated by the ANN based forecaster in which a novel contribution is taken in direct for altering the structure and parameters of the filter. The projected technique is based on the preceding delivery assumptions. Predictions are obtained by weighting up all probable models and restriction values according to their posterior distribution. Moreover, if the time series is soft or rough, the fitting algorithm can be transformed to suit, in utility of the extended or small terms to chiastic dependence of the time series, an on-line heuristic law to set the training process, transform the NN topology, adjust the number of patterns and iterations in calculation to the Bayesian inference in agreement with Hurst parameter H taking into description that the series forecasted has the same H as the real time series.

A. Geetha, G.M Nasira et al,[2013],Forecast is a challenging assignment and that too for weather is yet more complex, dynamic and mind-boggling. Weather forecast poses right from the earliest period as a big phenomenal task, since it depends on a variety of parameters to expect the dependent relative variables similar to temperature, rainfall, humidity, wind speed and direction, which are varying from time to time and weather estimate varies with the geographical place beside with its atmospheric variables. There are numerous data mining techniques employed for weather forecast, but choice tree evaluation can be quantified. This work places of interest a model using decision tree to forecast weather phenomena similar to fog, rainfall, cyclones and thunderstorms, which can be a lifesaving in sequence and worn by popular of all walks of life in manufacture prudent and sharp decisions. This model might be used in machine learning and further promises the extent for development as much more related attributes can be used in predicting the dependent variables. The model is implemented by means of the open source data mining tool Weka.[6]

Zhangjiajie, Hunan China et al [2013], In today's world there is sample occasion to cloud the frequent sources of time series data obtainable for decision making. This time efficient data can be used to pick up decision making if the data is transformed to in order and then into familiarity which is called knowledge discovery. Data Mining (DM) methods are being more and ore used in prediction with time series data, in accumulation to traditional statistical approaches. This work introduced an narrative review of the use of DM and statistical approaches with time series data, focusing on weather calculation. This is an part that has been attracting a huge transaction of notice from researcher in the pasture

3. DATA MINING ALGORITHMS

Research on data mining has led to the formulation of several data mining algorithms. These algorithms can be directly used on a dataset for creating some models or to draw vital conclusions and inferences from that dataset. Some popular data mining algorithms are SVM, Naïve Bayes, aritificial neural netork, etc. They are discussed in the follows section.

Support Vector Machine

SVM is a supervised Machine Learning algorithm which can be used for both classification and regression challenges. However it is mostly used in classification problems.SVM are a subclass of supervised classifiers that attempt to partition a feature space into two or more groups. Then, we perform classification by finding the hyper-plane that differentiates the two classes.

Support Vector Machines are particularly suited to handle such tasks. Support Vector Machine (SVM) is primarily a classier method that performs classification tasks by constructing hyper planes in a multidimensional space that separates cases of different class labels. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables.

Naive Bayes

Navie Bayes classifiers are highly scalable, requiring a number of parameter leaner in the number of variables (feature\ predictors) in a learning problem. It is a simple technique for constructing classifiers that assign. Class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. Navie Bayes classifier assumes that the value of a particular feature is independent of the value of any other feature, given the class variable .An advantages of Navie Bayes is that it only requires a small number of training data to estimate the parameters necessary for classification.

J48

J48 is an extension of ID3.The additional features of J48 are accounting for missing values, decision tress pruning, continuous attribute value ranges, derivation of rules, etc. In the WEKA data mining tool, J48 is an open source Java implementation of the C4.5 algorithm. After the 64byte protocol structure standardization and Genetic approach functions of mutation and cross over, the fitness of the protocol device identification is carried out, using the modified J48 decision tree algorithm.

Artificial Neural Network

ANN are types of computer architecture inspired by biological neural networks (Nervous system of the brain) and are used to approximate functions that can depend on a large number of inputs and are generally unknown .Artificial neural networks are presented as systems of interconnected "neurons" which can complete values from inputs and are capable of machine learning as well as pattern recognition due their adaptive nature. An artificial neural network achieves by creating connections between many different processing elements each corresponding to a single neuron in a biological brain. These neurons may be absolutely constructed or simulated by a digital computer system. Each neuron takes many input signals then based on an internal weighting produces a single output signal that is sent as input to another neuron.

Multi Layer Preceptor

A Multi Layer Preceptor Process (MLP-Model) is a feed forward artificial neural network model that maps set of input data onto set of appropriate outputs. An MLP model consists of multiple layers of modes in a directed graph, with each layer fully connected to the next one.MLP is a class of feed forward artificial neural network. An MLP consists of, at least, three layers of nodes: an input layer a hidden layer and an output layer.

Except for the input nodes each node is a neuron that uses a nonlinear activation function.MLP utilizes a supervised learning technique called back propagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable. A MLP can have more than one output neuron. The number of output neurons depends on the way the target values (desired values) of the training patterns are described.

4. METHODOLOGY

Step 1: Take the Rain Fall Dataset to Clustering Variables and greatest Number of groups (K Means Clustering)

Step 2: Introduce and Initialize cluster methods.

Example, value of K is measured as 13. Cluster centroids are implemented with first clarification.

Step 3: To compute Euclidean Distance

Euclidean is one of the distance method use on K Means algorithm. Euclidean distance among of a examination and original cluster centroids 1 to 13 is considered. Based on Euclidean distance all observation is assign to one of the clusters based on least amount distance.

Step 4: Take away Un similar Attributes on both train and test dataset. Accidentally initialize the weights in the network.

Step 5: To apply the input to the network and gather the computed output.

Step 6: Estimate the fault (e) e = desired-computed.

Step 7: Analyse the Δwi for all weights in rearward pass from hidden layer to output layer.

Step 8: Determine the ΔWi for all weights in rearward pass from input layer to hidden layer. Renew the weights in the network.

Step 9: Again the step 4 to 8 for each training model until all model classified accurately.

5. RESULTS AND DISCUSSION

SVM, AdaSVM, Naive Bayes, AdaNaive are the classification method used for time series predict in this research work. Two group are separated from the data set for training and for testing the algorithms of classification. To execute the classification algorithms, the tool used is Weka data examination. For classification procedure no more than a separation of data is particular from the loaded data. To choose a subset from innovative data, "Select attribute" are utilised by the operative. The preferred subset is then subjected to "X-Validation" operator. It develop the classification representation which is validated by the test data.

AdaBoost based SVM (AdaSVM), SVM, AdaBoost based Naive Bayes (AdaNaive) and Naive Bayes are implement for classification by using "X-Validation" operator. The presentation of the classification algorithm is evaluated by with the presentation operator. Performance estimate achieve for both the classification algorithms.

The HYBRID IMLP (Improved Multi Layer Perceptron), time lag recurring network and self organizing feature map are qualified for multi step in advance predictions and the results are compared with orientation to MSE (Mean square error),and NMSE (Normalized Mean Square Error) on testing in addition to training data set for short term prediction ,the number of experiments are approved by shifting a variety of parameters like no. of possessing elements ,number of hidden layers, no. of iterations, relocate function learning rule.

On the Rainfall time series number of experiments is approved out for the multi step ahead calculation. At this time the varying number of allowance essentials and effect out the number of hand out fundamentals for which the network give smallest amount MSE(Mean Square Error) .

Anywhere the accurate positive cases are denoted by TP, accurate negative cases are denoted by TN, FP and FN are denoted for phony positive cases and phony negative cases correspondingly.

The classification Error is $E_{t=(\frac{f}{n})} * 100$ where t represents the method, F denotes number of substance classified the wrong way and N reveals total number of samples. The presentation assessment achieved for together the categorization algorithms (existing and proposed) are given in Table 1.

Measures	Precision	Recall
AdaSVM	0.17	0.216
AdaNaive	0.522	0.545
J48	0.106	0.192
Hybrid IMLP(Proposed)	0.773	0.772

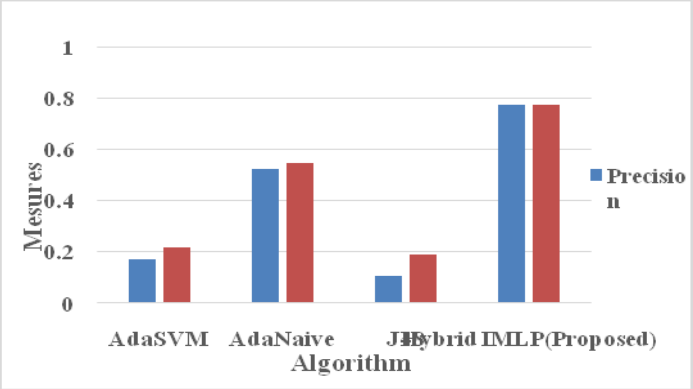

Fig.5.1 Precision and Recall value Comparison

Measures	Classification Accuracy(%)
AdaSVM	25
AdaNaive	27
J48	25
Hybrid IMLP(Proposed)	38

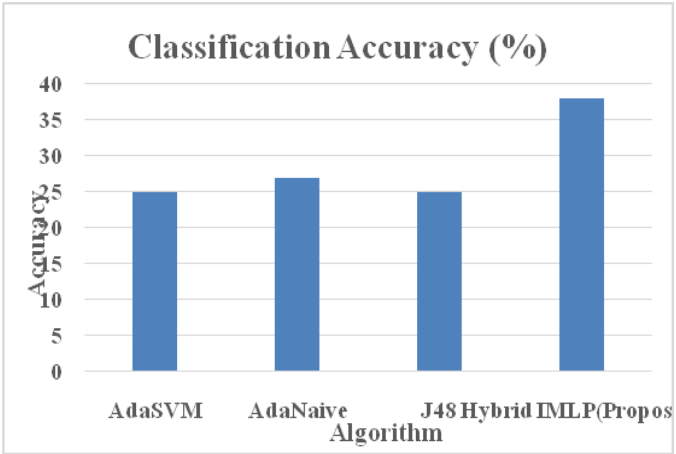

Fig.5.2 Accuracy Comparison

Measures	Execution time
AdaSVM	364
AdaNaive	100
J48	169
Hybrid IMLP(Proposed)	17

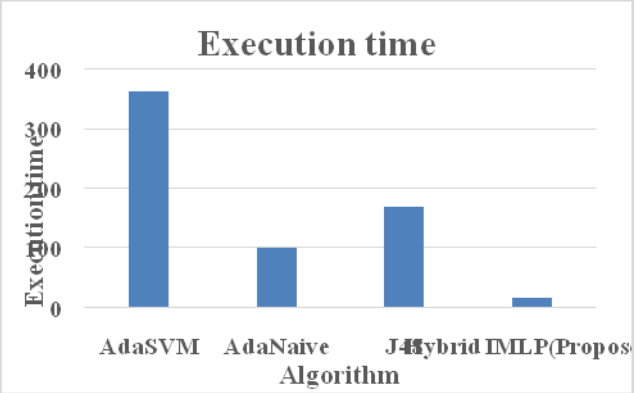

Fig.5.3 Execution Time

CONCLUSIONS

This research has been obtained a deep learning method based on the use of auto encoder and neural networks to forecast the accumulated rainfall for the next day. This approach forecasts the everyday accumulated rainfall in an exact meteorological position located in a middle area of India city. The results recommend that the proposed architecture better other approaches in conditions of the MSE and the RMSE.

It might be done from the results that there is good amount of augment in the accurateness of forecast and considerable reduce in the percentage of classification error in both the projected techniques called superior Multilayer Perceptron. Significance of rainfall forecast cannot be over emphasize. Permanent research for beating at mainly satisfactory method of prediction is essential.

REFERENCES

1. Cheng Zhou, Boris Cule, Bart Goethals "Pattern Based Sequence Classification", IEEE Transactions on Knowledge and Data Engineering, Vol. 28, 5, pp.1285-1298, 2016.
2. B. Tang, H. He, P. M. Baggenstoss and S. Kay "A Bayesian Classification Approach Using Class-Specific Features for Text Categorization", IEEE Transactions on Knowledge and Data Engineering, Vol.28, 6, pp.1602-1606, 2016.
3. P.Samuel Quinan, Miriah Meyer "Visually Comparing Weather Features in Forecasts", IEEE Transactions on Visualization and Computer Graphics, Vol. 22, 1, pp. 389-398, 2016.
4. Suhartono, Ria Faulina, Dwi Ayu Lusia, Bambang W. Otok, Sutikno, Heri Kuswanto "Ensemble Method based on ANFIS-ARIMA for Rainfall Prediction", IEEE International conference on statistics in Science, Business and Engineering (ICSSBE), pp.1-4, 2012.
5. C. R. Rivero, J. Pucheta, S. Laboret, M. Herrera and V. Sauchelli "Time Series Forecasting Using Bayesian Method: Application to Cumulative Rainfall", IEEE latin america transactions, Vol. 11, 1, pp. 359-364, 2013.
6. Jiawei Han And Micheline Kamber "Data Mining Concepts And Techniques", Second Edition 2006.,
7. Zdravko Markov, "An Introduction to the WEKA Data Mining System", 2006.
8. Alexandre Kowalczyk, "Support vector machines succinctly" 2000,.
9. Lorenzo Rosasco, "SVMC An introduction to Support Vector Machines Classification" 2012.
10. Deepti Gupta, Udayan Ghose, " A Comparative Study of Classification Algorithms for Forecasting Rainfall ", IEEE 4th International conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), pp. 1-6, 2015.