Exploratory Analysis of Rain Fall Data in India for Agriculture

In India, Agriculture is directly depends on the rainfall. The Indian economy is mainly based on the growth and prosperity of the agriculture sector. This entire agriculture field relies on rainfall due to very less and improper arrangement of irrigation and sources of water. Since the last few years, the pattern of rainfall has been changed in the entire region and it is unpredictable. The accurate prediction of rainfall may help in deciding by the farmers for which lots of techniques are available ranging from MLR, Artificial neural networks, etc.

It is important to estimate accurate rainfall for the effective use of water resources and optimal planning of water structures and availability. The various models and techniques are developed to estimate rainfall in various researches using data mining techniques. The accurate estimation of rainfall prediction and estimation of precipitation is not possible though many techniques are available. The use of data mining techniques to predict rainfall and its consequences may prove significant in the prediction of accurate rainfall that will help in the growth of the agriculture sector and the farmers can take their decisions accordingly. This paper studies various techniques of rainfall prediction and estimation and their results with the actual rainfall value.

LITERATURE SURVEY

There are different techniques used for the prediction of rainfall such as Regression analysis, clustering, and Artificial Neural Networks (ANN). Fundamentally, two approaches are used for predicting rainfall. One is the Empirical approach and the other is Dynamical approach. The empirical approach is based on an analysis of historical data of the rainfall and its relationship to a variety of atmospheric and oceanic variables over different parts of the world. The most widely used empirical approaches used for climate predictions are regression, artificial neural network, fuzzy logic, and group method of data handling. In Dynamical approach, predictions are generated by physical models based on systems of equations that predict the evolution of the global climate system in response to initial atmospheric conditions.

1. Jyothis Joseph described the empirical method technique belonging to the clustering and classification approach. ANNs are used to implement these techniques. Relative Humidity, Pressure, Temperature, Precipitable Water, Wind Speed. subtractive clustering is used. Subtractive clustering is a fast, one-pass algorithm for estimating the number of clusters and the cluster centres in a set of data. Applying subtractive clustering, the optimum numbers of clusters are obtained. The rainfall values are categorized as low, medium & heavy. The classifier model has been evaluated using a confusion matrix and the results have been obtained. The neural network Bayesian regularization has been applied in the

implementation. There is two methods such as classification and clustering are implemented.

2. Wint Thida Zaw, Thinn Thu Naing proposed the modelling of monthly rainfall prediction over Myanmar by applying the polynomial regression equation. The statistical relationship between rainfall amount and other climatic data is searched with the use of second-order MPR equation which contains added terms and nonlinear cross-product interaction of n predictors expressing with the first and second power of the predictors. Then the predictors which have high intercorrelation with others are reduced because the presence of many highly intercorrelated explanatory variables may substantially increase the sampling variation of the regression coefficients, and not improve, or even worsen the models predictive ability. Experiments and graphs are reported for Pathein rain gauge station located in lower Myanmar and Magwe station located in upper Myanmar where rainfall prediction is needed more for agriculture planning and management.

3. Narasimha Prasad proposes a need for the models for improving accuracy in the precipitation prediction using the Supervised Learning in Quest (SLIQ) decision tree using the Gini index for the prediction of the precipitation. To employ the SLIQ decision tree using a gain ratio that improves the accuracy using attributes such as humidity, temperature, pressure, wind speed and dew point. For every attribute, they found a split point using the attribute and its corresponding

class label pair wherever there is a change in the class label. For every split, a point identified to find the midpoint for the changed class labels and proceed until it reaches the end of the data. Compare all the split points gain ratio values and the maximum value is the best split point for that attribute. The gain value obtained for the attribute is to be divided by split info value of the class label, to obtain the gain ratio value for that attribute as shown in equation

Gain Ratio (V) = Gain(V) / Split Info(V).