International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 24 Issue 10 – JULY 2017.

A Determined Way for Crop Prediction by Using Pam and Decision Tree Algorithms

Ningappa Uppal [#1] and Dr Arun Biradar [*2]

[#] *Department of CSE,EWIT, Bangalore, Karnataka, India.*

[*] *Department of CSE,EWIT, Bangalore, Karnataka, India.*

***Abstract*— Prediction of Crop is an very important factor for agriculture and farmers. India is one of the agricultural countries; Crop forecast is an big challenging problem, so wide variety of Crop forecast methods are used. In this work particularly Karnataka and Tamilnadu states with some districts data are collected namely from Chitradurga, Mangalore, Bangalore, Gadag, Belgam and Mysore. From Tamilnadu one district data is collected namely triuchirapalli. From past 11 years (2005-2016) required data are collected with many parameters like wind speed, sea level pressure, min temperature, max temperature, mean temperature, dew point, precipitation. For prediction of Crop , methods like K means and artificial neural network algorithms are used in existing system for clustering and classification. In proposed system farthest first algorithm is used for clustering and j48 algorithm is used for classification for Crop forecasting. The collected sample data is clustered according to feature of farthest first algorithm by calculating distance between the points using Euclidian formula. Then the clustered data are then passed for training and testing. In this tool like eclipse mars .2 and also java programming language is used. The results obtained from proposed algorithms like farthest first and j48 gives better accuracy for Crop forecast and compared this result with existing algorithms.**

***Index Terms*— Crop prediction, clustering, classification, farthest first, j48.**

I. INTRODUCTION

The major challenges for agriculture and flood management can be overcome by using Crop prediction. Over the past years many attempts and methods have been used for Crop forecast. The prediction of Crop can be done by using historical data.

India is one of the agricultural country so for this prediction of Crop is very important. It is made by collecting required and quantitative data, here perticularly karnataka and tamilnadu with specific districts data is collected to predict Crop for particular regions like in Karnataka: chitradurga, mysore, mangalore, bangalore, gadag, and belgaum, in Tamilnadu: Triuchirapalli. Past years data is collected from 2005-2016 years.

Crop is not static it is dynamically changes from year to year depending on parameters, they are temperature, humidity, wind speed etc. In this parameters like min temperature, max temperature, mean temperature, wind speed, dew point precipitation, sea level pressure all these attributes are considered for Crop forecast.

The data mining techniques called clustering and classification are used. Farthest first algorithm is used to make different clusters for collected data often it will use full for analysis and gives good performance. Then for classification j48 algorithm is used.

II. LITERATURE REVIEW

Many researchers have tried to obtain the Crop with their respective data mining techniques. Crop forecast with timely and accurately is an very big challenging task. Generally for prediction the unsupervised algorithms suites very well to make clustering and supervised algorithms suites for classification.

Pinky saikia dutta [1], Hitesh tahbilder were tried for prediction of Crop model based on Multiple linear regression indicates acceptable accuracy F test, T test is used for explanatory variables. In this forecasting of monthly Crop by using data mining techniques over assam have been done. The six years of period[2007-2012] of data is collected from Regional Meteorological Center. Some of parameters like maximum temperature, minimum temperature, wind speed and Mean sea level have been considered for prediction. wind direction is not considered which could give more accurate result.

P.Hemalatha [2], used the data mining technique for guiding the path of the ships. To identify where the ship is navigating GPS(Global Positioning System) is used. The weather report of the ships location is made to traverse the Farthest-first and the corresponding decision is passed to the ship for its safe navigation by using farthest-first .

Badhiye S. S., Dr. Chatur P. N., Wakode B. V [3], were tried to predict the condition of climate with specific and very important attributes or parameters like temperature and humidity. The main aim of these two authors were to acquire temperature and humidity data and use a clustering technique with k-Nearest Neighbor method to find the hidden patterns inside the large dataset so as to transfer the retrieved information into usable knowledge for classification and prediction of climate condition.

Neha Khandelwal and Ruchi Davey [4], were first tried to consider weather daily historical data collected at JAIPUR city for trying to extract useful knowledge. Then data mining

technique called regression analysis is applied for Crop forecasting by using different four climate factors temperature, humidity, pressure, sea level.

With the combination of k means clustering and artificial neural network the Akash Dutt Dubey [5], have done prediction of Crop very accurately. The data is collected from period 1951-2010 with parameters like monthly relative humidity, atmospheric pressure, wind speed and average temperature. The result is compared from Regression results on RBFNN Training with Regression results on Kmeans based RBFNN Training which gives very good accuracy.

Sam Cramer, Michael Kampouridism, Alex A. Freitas [6], were used the feature engineering for improving financial derivatives based Crop prediction. In this the GP(genetic programming) using feature engineering on 21 different data sets of cities around Europe. The length of data is considered to be 10 years of daily Crop for training and one year of daily Crop for testing. Comparing to MCRP, STR, Crop runoff, the GP works well and overcomes the disadvantages of previous methods. By using GP with the hourly data prediction can be done accurately without any assumptions and it supports non linear data also.

Kavita Pabreja, [8], forecast the cloud burst by using clustering technique which helps to get the numerical weather prediction(NWP) for temperature and humidity. The important parameters required to cloud burst are temperature and humidity.

Amr H. El-Shafie1, A. El-Shafie2, Hasan G. El Mazoghi2, A. Shehata3 and Mohd. R. Taha1 [7], used ANN and MLR algorithms for Crop prediction over Alexandria, Egypt area. The data set used in this includes daily measurements for Crop and temperature from period 1957-2009.

III. Existing System

In existing work ANN algorithm is used to predict the Crop with K means clustering algorithm. With these two combinations good accurate result was obtained, but ANN have two issues which limits its application, the computation complexity of the network and the learning time. In order to overcome these two problems k means clustering algorithm is used for this work.

In this data is collected from one of the Japanese city called Yokohama from period of year 1951-2010. The parameters used in this to forecast the Crop are temperature, humidity, wind speed, atmospheric pressure. The main disadvantage in this work is if ANN is using means, K means is compulsory to overcome ANN issues.

IV. Datasets Used

With different combination of techniques Crop forecast is done in this proposed work. So for this data is collected from period of year 1998-2009, with particular region of states in India as Karnataka and Tamilnadu. Specific districts data is collected from Karnataka are Chitradurga, Mangalore, Mysore, Bangalore etc.. in Tamilnadu are Triuchirapalli. Particular attributes are considered for this work as min temperature, max temperature, mean temperature, wind speed, sea level pressure, dew point and precipitation.

V. Techniques Used

A. *Kmeans algorithm*

The Decision Tree traversal of a finite point set may be computed by greedy algorithm that maintains the distance of each point from the previously selected points, performing the following steps:

STEP 1: Initialize the sequence of selected points to the empty sequence, and the distances of each point to the selected points to infinity.

STEP 2: While not all points have been selected, repeat the following steps:

STEP 2.1: Scan the list of not-yet-selected points to find a point *p* that has the maximum distance from the selected points.

STEP 2.2: Remove *p* from the not-yet-selected points and add it to the end of the sequence of selected points.

STEP 2.3: For each remaining not-yet-selected point *q*, replace the distance stored for *q* by the minimum of its old value and the distance from *p* to *q*. Farthest-point traversals have many applications, as same as of the traveling salesman problem and the K center problem. K means(Data,k) Centers<--the set consisting of a single randomly chosen point fromDat while|Centers|<k DataPoint <--the pointin Data maximizing d(DataPoint,Centers) add DataPoint to Centers return Centers

B. *J48 algorithm*

1) Association rules:

Used to obtain integrating relations between variables in large database.

2) Rules evaluation metrics:

- support(A+B)=No. of transactions containing A&B No. of total transactions
- Confidence(A+B)=No. of Transactions containing A&B No. of Transactions containing A

3) Pseudo code of j48 algorithm:

Step1: CK: Candidate item set of size k.**Step2: LK:** Frequent item set of size k.

Step3:LT=(frequent items); for (k=1;LK !=0;k++) do begin

Step4:CK+1=candidates generated from LK;

Step5: for each transaction t in database do.

Step6: increment count of all candidates in CK+1 that are contained in t.

Step7:LK+1=candidate in CK+1 with min support.

Step8: end

Step9: return kLK;

4) J48 algorithm flow:

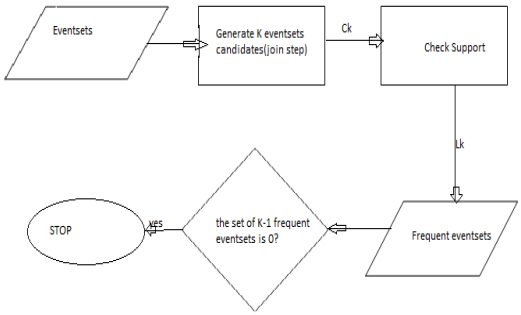

Fig5.3.1: J48 algorithm flow

The above figure shows the j48 algorithm flow. There are two steps to perform j48 i) Join step and ii) prune. The entire algorithm is classified into two steps as follows:

STEP 1: Apply min support to find all the frequent event sets with k items in a database.

STEP 2: Use the self-join rule to find the frequent event sets with k+1 items with the help of frequent k-events. Repeat this process from k=1 to the point when we are unable to apply the self-join rule. By applying farthest first, clustering have been done for analysis and after this j48 algorithm is used to make classification and by applying with these two combination of algorithm better accuracy can be obtained compare to existing combination of algorithms.

VI. RESULTS AND DISCUSSION:

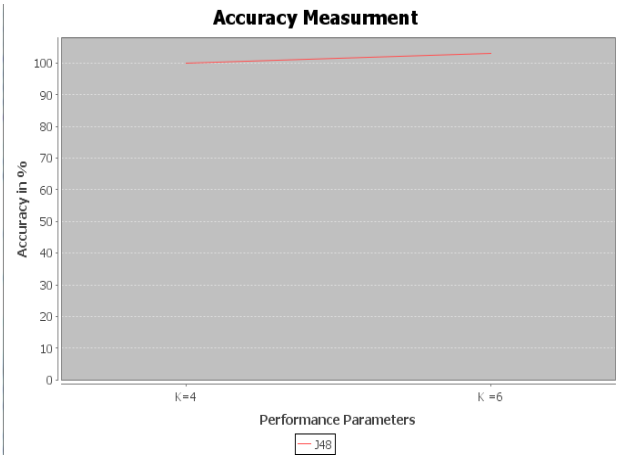

Fig a. Classification Accuracy

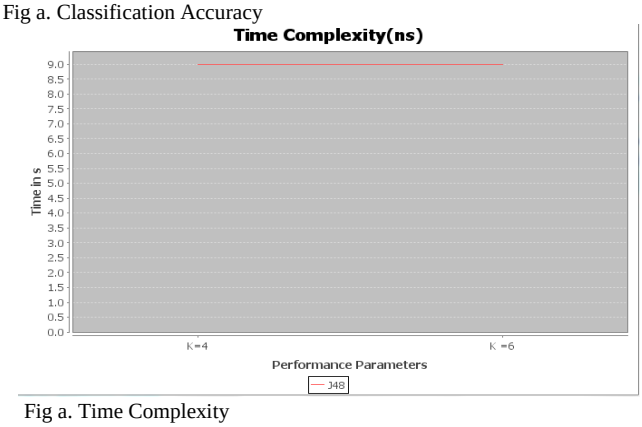

Fig a. Time Complexity

Above fig(a), depicts the classification accuracy of crop prediction data set around 98% is achieved using proposed techniques.

Similarly from fig(b), it has been predicted that for crop prediction in Karnataka state, proposed algorithm performance quickly wuth a time period of just 8.5s.

VII. CONCLUSION

In this paper working principle PAM and J48 combination of algorithms are proposed for Crop prediction. The data is collected from Karnataka for specific district regions. The data samples were first clustered into different clusters according to feature of PAM algorithm and then classification have been made by using J48 algorithm. By using these two combinations better accurate result can be obtained compared to k means with ANN, finally proposed technique with next year prediction for 10000 Hectares is achieved the high accuracy its around 98%, when compared with other techniques, for the future supposed to work with crop datasets with multiple states,

REFERENCES

[1] Pinky saikia dutta, Hitesh tahbilder, "Prediction of Crop using data mining technique over Assam", Apr-May 2014. Indian Journal of Computer Science and Engineering.

[2] P Hemalatha, "Implementation of data mining technique for weather report guidance for ships using GPS", march 2013, International Journal Of Computational Engineering Research.

[3] Badhiye S. S., Dr. Chatur P. N., Wakode B. V, "Temperature and humidity data analysis for future value prediction using clustering technique", 1, January 2012. International Journal of Emerging Technology and Advanced Engineering.

[4] Neha Khandelwal and Ruchi Davey, "Climatic Assessment Of Rajasthan's Region For Drought With Concern Of Data Mining Techniques", September- October 2012. International Journal of Emerging Technology and Advanced Engineering.

[5] Akash Dutt Dubey, ": K-Means Based Radial Basis Function Neural Networks for Crop Prediction", 2015.IEEE paper.

[6] Sam Cramer, Michael Kampouridism, Alex A. Freitas, "Feature Engineering for Improving Financial Derivatives-based Crop Prediction", 2016 IEEE Congress on Evolutionary Computation (CEC).

[7] Amr H. El-Shafie1, A. El-Shafie2, Hasan G. El Mazoghi2, A. Shehata3 and Mohd. R. Taha1, " Artificial neural network technique for Crop forecasting applied to Alexandria, Egypt", International Journal of the Physical Sciences Vol. 6(6), 1306-1316, 18 March, 2011.

[8] Kavita Pabreja, " Clustering technique to interpret Numerical Weather Prediction output products for forecast of Cloudburst", International Journal of Computer Science and Information Technologies, Vol. 3 (1) , 2012,