EFFICIENT CLUSTERING TECHNIQUE TO ANALYZE THE DROUGHT IN AGRICULTURE

S.Shivaprasad[1] Dr. U. Srilakshmi[2] G.Parimala[3]

[1,2,3] **Assistant professor, Department of CSE, VFSTR Deemed to be University, Guntur.**

Abstract:

Drought is one of social issue and severe problem faced by formers now days. We are focused to reduce the damage of droughts while cultivating the crops. It gives the better estimation to the formers to get required outcome of crop by analysing data based on crop productivity, amount of rainfall, agricultural inputs, irrigation, and similar factors for every crop. We created dataset that based on reports specifying these actions, studied and analysed from data taken over the past few years. We analyzed the Data to extract the relations in between total irrigation area and type of crop; total principal and non-principal crop amount versus district-wise rainfall etc. Actions are used to reduce the damage of drought, which will also be suggested as best crop. We applied the Hierarchical clustering to analyze the crops and it produces 78.3 % accuracy.

1. Introduction:

India is an Agriculture country, its economy mostly depends upon agriculture growth. Agriculture is largely influenced by rain water, which is highly unpredictable. In the present scenario, the government also providing but this data is not useful to many of the people like Farmers. This trend will act as solutions for farmers, especially in drought affected area for example the cultivation of Kharif crop is suitable mainly in rainy season, otherwise we can use the ground water resources but now-a-days the amount of ground water was reduced, which consequently leads to drought like conditions. They are some other example like maize, which was sown in winter and harvested in spring season.Rabi crops are sown at mid November and whenever the monsoon rain was completed.But the harvesting was started at April or May.

We are cultivating the crops with the help of rain water which was stored.If there is more amount of rainfall in winter season then it is profit to kharif crop but at the same time it as the disadvantage for Rabi crop. If you look at these examples, you will see that the amount of rainfall that occurs in an area, the crop produced in that area, the season and other such parameters are all correlated and can form trends that can give us solutions for suitable farming practices that can help farmers successfully avoid drought-like situations.

2. LITERATURE SURVEY

Agricultural growth also depends on the soil fertility namely nitrogen, phosphorus, potassium etc.[3][4][5] Now a day's India is going towards the development of technical side which is beneficial for the agricultural land which will increase the productivity of the crop.

This results the better productivity for the farmer for their corresponding agricultural field. The farmer suicides, which have remained unstoppable for past few years in India. The practice of farming is one of the major occupations in our country, and a major produce of a variety of crops come from the state of Andhra Pradesh. Information Mining is an inclining research field in horticultural harvest yield examination. In this undertaking, our emphasis is on the uses of Data Mining Techniques in the horticultural field.

Various Data Mining methods are being used, for example, K-Means, K-Nearest Neighbour (KNN) and Support Vector Machines (SVM) for various sorts of uses in Data Mining strategies for the horticultural field [5]. In this venture, we can consider the issue of foreseeing drought. Drought is a perfect horticultural issue that remaining parts to be explained dependent on the accessible information.

The issue of yield forecast can be comprehended by applying Data Mining procedures [7]. This procedure targets finding appropriate information models that get exceptionally precise outcomes and a high sweeping statement in the type of yield expectation abilities. For this reason, various kinds of Data Mining strategies were assessed on various informational collections. The administration is gathering information just in its crude structure, and this information is of no utilization to the end-client, that is the formers. Gathering this information, normalizing it, dissecting it, that will give the extent of our undertaking. In this work, the exhibition of a promising framework that is just planned for helping ranchers in dry season influenced regions.

This information would first be able to be normalized and it was broke down to discover different relations that can assist with finding the answers for ranchers. Droughts affect our farmers the most, as is made evident by the struggle faced by number of farmers' suicide cases in the state of Andhrapradesh due to droughts. Thus, this issue must be atmost importance to the government. To help reduce the sufferings of farmers, governments of various states and local bodies of a number of districts can make use of this system to generate reports and find solutions for their farmers.

3. Proposed Solution:

In our work there are two phases in the proposed solution i.e training and testing. The database created for the analysis using records for rainfall, temperature and pressure for some districts of India. For the training 80% of the data will be used for training the system and the 20% of the data will be used for testing the accuracy of the system.

A. Phase One

In phase one of the system comprises of extraction of data related to parameters required for drought classification, which are rainfall, temperature and pressure, fertility, number of labour, Season, crop. This data will then be standardized to have a consistent format and loaded into the database.

The analysis of this data will produce results with other parameters, which can be used to calculate the probability of drought and classifying it as low, medium or high.

Steps followed in phase one:

1. Extract the data from available places,

2. Transform the data into a standard format and then load into the database.

2.1 Analysing the Rainfall data,

2.2 Analysing the Temperature data,

2.3 Analysing the Pressure data,

2.4Analysing the Fertility data,

2.5Analysing the Number of labours present,

2.6Analysing the Season,

2.7Analysing the Crop data.

3. Correlation analysis for drought prediction.

4. End the Phase 1.

Phase Two:

The phase two of the system will obtain the result and attributes from phase one and feed it into the next classifier. Along with that, data regarding other parameters like soil type, fertilizer input etc. will also be correlated to find the Crop Growth Probability Index which will be mentioned in the form of a report

1. Start the phase2.

2. Obtain the drought prediction from phase1.

3. Correlate data for seasonal parameter.

   Correlate fertilizer data and statistics.

   Corelate data for type of crop.

   Corelate soil data and statistics.

4. Generate reports for users.

5. End of phase 2.

It is a brief survey of data mining techniques that can be used, and have been used in the past for agricultural datasets. It provides an overview of the following data mining techniques – Classification, Clustering, Association Rule Mining, and Regression. It goes onto illustrate several examples for these algorithms, one of them being as follows. Support Vector Machines can be used in case of crop classification, in case of changing weather conditions.

3.1 Data Collection:

We are collected information from different formers belongs to different regions. They expressed different attributes regarding droughts in the crops and reaming problem. As per our project we are consider the some of the attributes to work on droughts. We are collected data in perspective of crop and the attribute we are collected as Temperature, Rainfall, fertility and area in acres, Season, profits and required labour. Data can originate from different sources, and needs to be checked before it can be put to use.

```
In [7]: dmt.head(3)
```

Out[7]:

	place	year	crop	Temperature	Rainfall	fertility	area in acrs	Season	profit	required labour
0	Vijayawada	2001	Wheat	30	high	good	2.0	whole year	1,50,000	10
1	Vijayawada	2002	Wheat	32	low	bad	3.0	whole year	1,70,000	15
2	Vijayawada	2003	Wheat	28	medium	bad	4.0	whole year	2,50,000	20

```
In [6]: dmt.tail()
```

Out[6]:

	place	year	crop	Temperature	Rainfall	fertility	area in acrs	Season	profit	required labour
55	nuziveedu	2001	paddy	27	high	good	2.0	kharif	1,40,000	10
56	nuziveedu	2002	paddy	22	low	bad	3.0	kharif	80000	15
57	nuziveedu	2003	paddy	23	high	average	6.0	kharif	150000	20
58	nuziveedu	2004	paddy	28	medium	good	5.0	kharif	300000	18
59	nuziveedu	2005	paddy	26	low	bad	7.0	kharif	300000	25

Fig.1 Descripition of database created

This can be done by directly importing files that may already be available in .csv or .xlsx formats.

3.2 Data Pre-processing:

Datasets in any data mining applications can have missing data values. These missing values can get propagated due to lack of communication among the parameters in a data collection system. These missing values can affect the performance of a data mining system, and it should be noticed.

```
In [30]: dmt.describe()
```

Out[30]:

	year	Temperature	area in acrs	required labour
count	60.000000	60.000000	59.000000	60.000000
mean	2003.000000	27.966667	4.338983	17.700000
std	1.426148	4.430123	1.908332	6.505018
min	2001.000000	21.000000	1.000000	5.000000
25%	2002.000000	24.000000	3.000000	15.000000
50%	2003.000000	28.000000	4.000000	17.500000
75%	2004.000000	31.250000	6.000000	23.250000
max	2005.000000	36.000000	8.000000	30.000000

Fig2. Statistical descrption of dataset created

```
In [16]: dmt.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60 entries, 0 to 59
Data columns (total 10 columns):
place                60 non-null object
year                 60 non-null int64
crop                 60 non-null object
Temperature          60 non-null int64
Rainfall             60 non-null object
fertility            59 non-null object
 area in acrs        59 non-null float64
Season               60 non-null object
profit               60 non-null object
required labour      60 non-null int64
dtypes: float64(1), int64(3), object(6)
memory usage: 4.8+ KB

In [17]: dmt.columns.tolist()
Out[17]: ['place',
          'year',
          'crop',
          'Temperature',
          'Rainfall',
          'fertility',
          ' area in acrs',
          'Season',
          'profit',
          'required labour']

In [18]: dmt.columns
Out[18]: Index(['place', 'year', 'crop', 'Temperature', 'Rainfall', 'fertility',
                ' area in acrs', 'Season', 'profit', 'required labour'],
               dtype='object')
```

Fig3. Different attributes of database

4. Proposed Methods

Clustering:

Clustering is a gathering of items that has a place with a similar class. Just, comparable items are assembled in one bunch and disparate articles are gathered in another group. While doing bunch investigation, we first segment the arrangement of information into bunches dependent on information comparability and ten dole out the names to the gatherings. We have various techniques for bunching. They are Partitioning Method, Hierarchal Method, Density-based Method, Grid-based Method, Model-based Method, and Constraint-based Method From these techniques; we have picked various level strategies for our venture.

4.1 Hierarchal Clustering:

This bunching includes making groups that have transcendent requesting through and through. This technique makes various level deterioration of the given arrangement of information objects. We can characterize progressive strategies based on how the various level decay is shaped. There are two methodologies here-

1. Agglomerative Approach,

2. DIVISIVE Approach.

Agglomerative Approach:

We are utilizing this methodology for our undertaking. This methodology is otherwise called the base up approach. In this, we start with each article shaping a different gathering. It continues doing as such until the entirety of the gatherings is converted into one or until the end condition holds. In this, the primary thought of this agglomerative grouping is to

guarantee the close by information focuses and end up in the same cluster. It starts with a collection of some clusters and it repeat until only one cluster is left. First we will find the pair of clusters which are closest based on the minimum distance. Next, Merge the individual clusters i.e. x, y into a group i.e., (x+y) which forms a new cluster.Now,we have to remove the individual clusters and add the new cluster in that place. Finally, it produces a DENDOGRAM-which is a hierarchical tree of clusters.We have to define the distance metric over the clusters.

4.2 APPLICATIONS OF DATAMINING IN AGRICULTURE:

In this modern times the technology is increasing drastically,To do correct type of farming we require more information in different aspects,to get accurate results.It will take more time to get such type of information.how number of labours required,temperature,rainfall,fertility etc.so,there are many aspects are required. Through datamining the agricultural organizations can give us predictive and descriptive information to us.As we know that agriculture is very complex.Today a good work to available the information to the farmers in electronic format.The key to get more productivity&profit is to extract the useful information for high quality of agriculture.

5. RESULTS:

```
In [22]: dmt["Temperature"].groupby(dmt["Temperature"], axis=0).count()
Out[22]: Temperature
         21      5
         22      3
         23      6
         24      3
         25      1
         26      2
         27      5
         28     11
         29      3
         30      3
         31      3
         32      6
         34      3
         35      3
         36      3
         Name: Temperature, dtype: int64
```

Fig4. Grouping of temparatutes

Agglomerative clustering

```
In [49]: from sklearn.cluster import AgglomerativeClustering
```

```
In [48]: hc = AgglomerativeClustering(n_clusters = 3, affinity = 'euclidean', linkage='ward')
         y_hc = hc.fit_predict(x)
         plt.scatter(x[y_hc == 0,0], x[y_hc == 0,1], s = 100, c ='cyan', label = 'Cluster 1')
         plt.scatter(x[y_hc == 1,0], x[y_hc == 1,1], s = 100, c ='red', label = 'Cluster 2')
         plt.scatter(x[y_hc == 2,0], x[y_hc == 2,1], s = 100, c ='blue', label = 'Cluster 3')
         plt.title('clusters of crops')
         plt.xlabel('Temperture')
         plt.ylabel('required labour')
         plt.legend()
         plt.show()
```

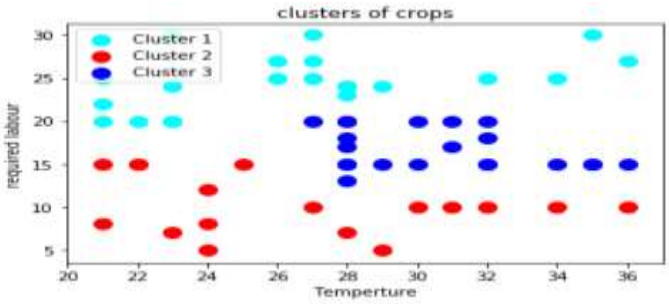

Fig.5 Scatter plot of database in between temparature and labour

```
In [50]: y = dataset.iloc[:, [1, 9]].values
```

```
In [51]: import scipy.cluster.hierarchy as sch
         dendrogram = sch.dendrogram(sch.linkage(y, method = 'ward'))
```

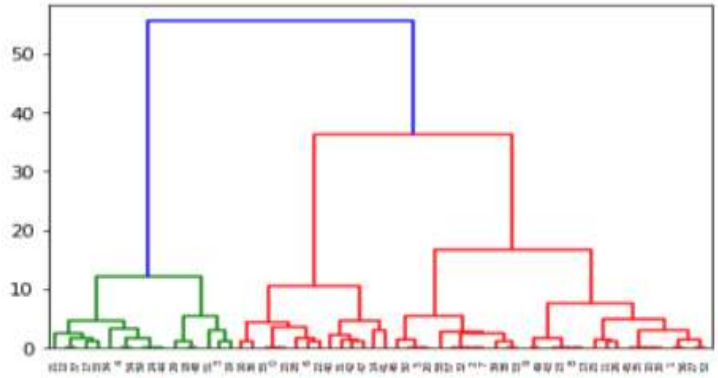

Fig.6 Dendogram plotting for Agglomerative cluster

```
In [52]: from sklearn.cluster import AgglomerativeClustering
hc = AgglomerativeClustering(n_clusters = 3, affinity = 'euclidean', linkage='ward')
y_hc = hc.fit_predict(y)
plt.scatter(y[y_hc == 0,0], y[y_hc == 0,1], s = 100, c ='green', label = 'Cluster 1')
plt.scatter(y[y_hc == 1,0], y[y_hc == 1,1], s = 100, c ='red', label = 'Cluster 2')
plt.scatter(y[y_hc == 2,0], y[y_hc == 2,1], s = 100, c ='black', label = 'Cluster 3')
plt.title('clusters')
plt.xlabel('year')
plt.ylabel('required labour')
plt.legend()
plt.show()
```

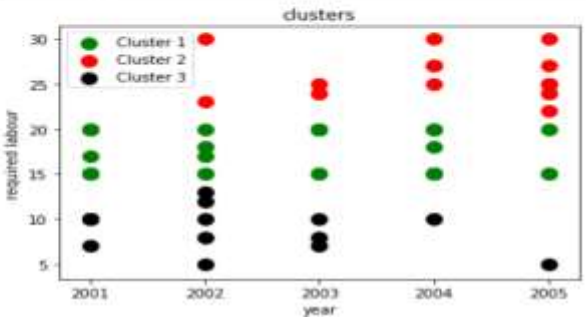

Fig.7 Scatter plot of database in between year and labour

```
In [53]: m = dataset.iloc[:, [1, 3]].values
         import scipy.cluster.hierarchy as sch
         dendrogram = sch.dendrogram(sch.linkage(m, method = 'ward'))
```

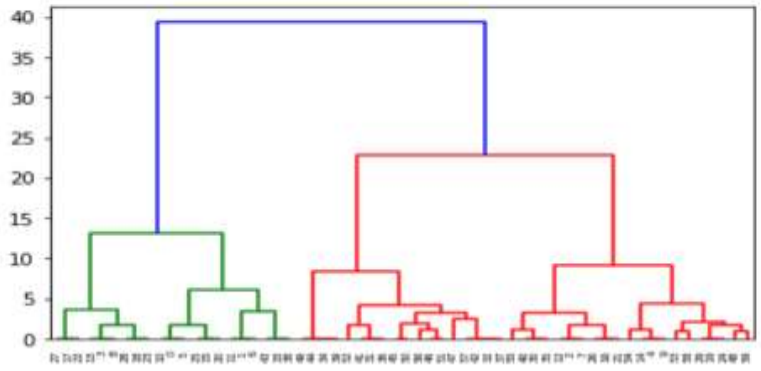

Fig.8 Dendogram in between year and labour

Accuracy

```
In [37]: from sklearn.metrics import accuracy_score
         print('Accuracy score: ', accuracy_score(y_test, y_pred))
```

```
Accuracy score:  0.7833333333333333

              precision    recall  f1-score   support

         0.0       0.53      0.64      0.58        14
         1.0       0.88      0.83      0.85        46

    accuracy                           0.78        60
   macro avg       0.71      0.73      0.72        60
weighted avg       0.80      0.78      0.79        60
```

Fig.9 Accuracy of model

6. Conclusion

We conclude that the application of data mining algorithms in the field of agriculture says about Crop productivity and drought predictions, if presented in a proper format to the end-users, the farmers, it will help drought affected villages and districts. Decision making based on analytical models will complement traditional methods of farming, leading to decreased costs and increase in crop yeild.This is used to make correct decisions in agriculture through data mining.There is large amount of data is available now a days,the farmer can be confused automatically,so to avoid such type of problems we use datamining.

Future Scope:

We want to add more attributes to it and to more accurate the results. Hence, after testing on a set of attributes, we can scopealso addother parameters such as agricultural inputs, soil nutrients and irrigated area. These parameters may improve the accuracy.Unsupervised clustering to label data for classifiers will also improve accuracy, instead of using fixed

intervals for the same. Then, after completing successfully , In further this system can be implemented in India and other countries , where to reduce the suffering of farmers.Each and every government is facing this issues , if we use this type of system we can find the solutions to the problems.

REFERENCES

1. Amiksha Ashok Patel " DATA MINING TRENDS IN AGRICULTURE : A REVIEW " AGRES – An International e. Journal Vol. 6, Issue 4:637-645, 2017.

2. Bhargavi, P, and Jyothi, S. (2009). Applying Naive Bayes data mining technique for classification of agricultural land soils. Int. J. Compt. Sci. Network Security, 9(8): 117-122.

3. Breiman, L.; Friedman, J. H.; Olshen, A. R. and Stone, C. J. " Classification and Regression Trees" 1982.

4. Soria-Olivas, E.; Martín-Guerrero, J. D. and Moreno, J. " Support vector machines for crop classification using hyper spectral data" Iberian Conference on Pattern Recognition and Image Analysis Pattern. pp. 134-141,2003.

5. Chi-Farn Chen; Ching-Yueh Chang; Jiun-Bin Chen "Spatiial knowledge discovery using spatial datamining method",Geoscience and Remote Sensing Symposium, IEEE International Volume 8, Issue 25, Page(s): 5602 - 5605 July 2005

6. Dr P Jaganathan , S.Vinothini & P.Backialakshmi "A Study of Data Mining Techniques to Agriculture" IJRIT International Journal of Research in Information Technology, Volume 2, Issue 4, April 2014, Pg: 306- 313.