EXPLORATORY ANALYSIS OF RAIN FALL DATA IN INDIA FOR AGRICULTURE

PROBLEM STATEMENT:

Rainfall is measured by a precise instrument called Rain gauze. It determines the depth of precipitation that occurs over a unit area. Thus, it measures rainfall. One liter of rainfall per meter squared is equivalent to one millimeter of measured precipitation.Now that we have understood how rainfall occurs we will now see what exactly is monsoon? Monsoon is traditionally defined as a seasonal reversal of wind accompanied by corresponding changes in precipitation.

-511319104063(TEAM LEADER)

India is an agricultural country and secondary agro based market will be steady with a good monsoon. The economic growth of each year depends on the amount of duration of monsoon rain, bad monsoon can lead to destruction of some crops, which may result in scarcity of some agricultural products which in turn can cause food inflation, insecurity and public unrest. In our analysis we are trying to understand the behavior of rainfall in India over the years, by months and different subdivisions. The Indian subcontinent heats up more as compared to the Indian ocean as the sun is directly over the landmass. This creates a low pressure over the Indian subcontinent and a relatively low pressure over the Indian ocean.So when the wind starts flowing from high pressure area to low pressure are i.e. from sea to land, it picks up the moisture from the sea and while entering the indian subcontinent it comes in contact with the high terrains and hence precipitation occurs.

-511319104042

In agriculture sector where farmers and agribusinesses have to make innumerable decisions every day and intricate complexities involves the various factors influencing them. An essential issue for agricultural planning intention is the accurate yield estimation for the numerous crops involved in the planning. Data mining techniques are necessary approach for accomplishing practical and effective solutions for this problem. Agriculture has been an obvious target for big data. Environmental conditions, variability in soil, input levels, combinations and commodity prices have made it all the more relevant for farmers to use information and get help to make critical farming decisions. This

paper focuses on the analysis of the agriculture data and finding optimal parameters to maximize the crop production using data mining techniques like PAM, CLARA, DBSCAN and Multiple Linear Regression. Mining the large amount of existing crop, soil and climatic data, and analysing new, non-experimental data optimizes the production and makes agriculture more resilient to climatic change.

-511319104046

Clustering is considered as an unsupervised classification process [4]. A large number of clustering algorithms have been developed for different purposes [4,5,6]. Clustering techniques can be categorised into Partitioning clustering, Hierarchical clustering, Density-based methods, Grid-based methods and Model based clustering methods.

Partitioning clustering algorithms, such as K-means, K-medoids PAM, CLARA and CLARANS assign objects into k (predefined cluster number) clusters, and iteratively reallocate objects to improve the quality of clustering results. Hierarchical clustering algorithms assign objects in tree structured clusters, i.e., a cluster can have data point's representatives of low level clusters [7]. The idea of Density-based clustering methods is that for each point of a cluster the neighbourhood of a given unit distance contains at least a minimum number of points, i.e. the density in the neighbourhood should reach some threshold. The idea of the density-based clustering algorithm is that, for each point of a cluster, the neighbourhood of a given unit distance has to contain at least a minimum number of points [8].

There are different forecasting methodologies developed and evaluated by the researchers all over the world in the field of agriculture. Some of such studies are: Researchers like Ramesh and Vishnu Vardhan are analysed the agriculture data for the years 1965–2009 in the district East Godavari of Andhra Pradesh, India. Rain fall data is clustered into 4 clusters by adopting the K means clustering method. Multiple linear regression (MLR) is the method used to model the linear relationship between a dependent variable and one or more independent variables. The dependent variable is rainfall and independent variables are year, area of sowing, production. Purpose of this work is to find suitable data models that achieve high accuracy and a high generality in terms of yield prediction capabilities

-511319104047