International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

AGRICULTURAL ANALYSIS USING DATA MINING TECHNIQUES

Pavithra B[1], Janani A[2], Jeyat Rathan R[3], Vijay Franklin J[4]

[1,2,3] *UG Student, Bannari Amman Institute of Technology, Sathyamangalam, India*

[4] *Asst. Professor, Bannari Amman Institute of Technology, Sathyamangalam, India.*

Abstract - *In agriculture sector wherever farmers and agri businesses need to build numberless choices daily and complicated complexities involves the varied factors influencing them. A necessary issue for agricultural designing intention is that the correct yield estimation for the various crops concerned within the designing. Data processing techniques are necessary approach for accomplishing sensible and effective solutions for this drawback. Python is employed as a face for analysing the agricultural knowledge set. Jupyter Notebook is that the data processing tool used to predict the crop production. The parameter includes within the dataset are precipitation, temperature, reference crop, evapotranspiration, area, production and yield for the season from January to December. The information mining techniques like K-Means, KNN, SVM, and theorem network formula wherever high accuracy is achieved. It additionally discusses on totally different data processing applications in determination the various agricultural issues. It integrates the work of assorted authors in one place therefore it's helpful for researchers to induce info of current situation of knowledge mining techniques.*

1. INTRODUCTION

Agriculture is that the major supply of the Indian Economy. Day by day, the population will increase. Therefore the demand of food will increase. To induce eliminate these things farmers, agricultural scientists, and researchers are attempting for higher crop yield. Agriculture is that the backbone of the Indian nation. In spite of the actual fact that enormous areas in Asian country are brought underneath irrigation, solely common fraction of the cropped half is irrigated. The productivity of agriculture is incredibly low. Therefore because the demand of food is increasing, the researchers, farmers, agricultural scientists and government are attempting to place additional effort and techniques for a lot of production. And as a result, the agricultural information will increase day by day because the volume of knowledge will increase, it needs involuntary manner for these information to be extracted once required. Still nowadays, an awfully few farmers are literally mistreatment the new strategies, tools and technique of farming for higher production. Data processing will be used for predicting the long run trends of agricultural processes. Data mining tasks will be classified into 2 categories: Descriptive data processing and prophetic data processing. Descriptive data processing tasks characterize the overall properties of the information within the information whereas prophetic data processing is employed to predict express values supported patterns determined from known results. Prediction involves mistreatment some variables or fields within the information to predict unknown or future values of different variables of interest. As so much as data processing technique is concern; within the most of cases prophetic data processing approach is employed. Prophetic data processing technique is employed to predict future crop, meteorology, pesticides and fertilizers to be used, revenue to be generated then on. The yield prediction downside will be resolved by using data processing techniques like K means that, K nearest neighbour (KNN), Artificial Neural Network and support vector machine (SVM). Analysis paper aims at finding appropriate information models that accomplish a high exactness and a high generality with regard to four parameters specifically rain, year, production and space of sowing. For this purpose, different kind of Mining techniques were evaluated on totally different data sets.

2. WORKING METHOD

The main techniques for data mining include K means rules, K nearest neighbour, Bayesian Network Algorithm and Support Vector machine. The different data mining techniques used for solving different agricultural problem has been mentioned.

A.K means clustering

K-Means clustering is one in every of the clustering technique that method a cluster of information points into low range of clusters. For instance, the things in an exceedingly mall are clustered in several classes (medium, large, X-L are classified because the size of the dress). This can be a qualitative technique of rending a bunch of information. A quantitative approach is employed to live distinctive characteristics of the merchandise. In k clusters, the amount of information points have to be compelled are to be divided. The aim of this technique is to assign a cluster to every information. K-means algorithmic rule aims to search out the clusters positions that minimize the space from the information points to the cluster.

B.KNN classifier(k nearest neighbour)

K-Nearest Neighbors is one in all the classification algorithms in Data Mining. It is otherwise known as supervised learning model and lazy algorithmic program due to instance learning. It is used for various applications like pattern recognition, data processing and intrusion detection. Implementation is straightforward for small information sets. The coaching information doesn't want any knowledge regarding the

structure of knowledge before the analysis. Downside of this classifier is searching for the closest neighbour for every sample. Tons of houses is required once the coaching information is giant. The gap between check information and also the coaching information ought to be calculated for each checking information. So, the testing wants a great deal of your time. There are 2 phases during this classifier: coaching phase: Save the examples, Prediction phase: Get the check instance and notice the coaching set.

C.Bayesian Network

A Bayesian network is associate acyclic graph consists of edges and nodes with directions during which every edge represents to a conditional dependency, and every node represents to a single variable quantity. The probabilistic graphical model that uses Bayesian abstract thought for computations square measure known as Bayesian networks. Bayesian network , that encodes a distribution lead (Crop Year, Crop Production, Crop Temperature, Crop Rainfall, Crop n), from a given coaching set. We are able to then use the ensuing model so given a group of attributes the classifier supported returns the label that maximizes the posterior chance (Crop Year, Crop Production). Note that, by causing classifiers during this manner, we tend to have a square measure addressing the most concern expressed within the introduction. Take away the bias by the independence assumptions embedded within the naïve Bayesian classifier. Network with a comparatively sensible MDL score that performs poorly as a classifier and to grasp the attainable discrepancy between sensible prophetical accuracy and sensible MDL score, we tend to should examine the MDL score. Recall that the log probability term in Equation two is that the one that measures the standard of the learned model, which D = denotes the coaching set.

D.SVM Classification(Support Vector Machine)

Support Vector Machine algorithmic rule is an distinguished information analysis methodology and it's used for classification and regression techniques. Here the information points are planned in n-dimensional house with the worth of specific characteristics because the value of a particular coordinate. The classification is completed by finding the hyper-plane line that separates the categories single. The procedures for using support vector machine algorithms are defined. Read the crop dataset. Produce the information frame and extraction feature of Crop production, year, temperature, mean downfall and mean temperature dataset. Produce the SVM category as e107 package and linear, non-linear and kernel sequences model. Declaration crop for temperature is the primary part scan information frame and set list of crop space and temperature. The second part applied support vector matrices to prediction state for crop dataset. The SVM results calculate on regression format dataset. Declaration crop for year of production is the primary part scan information frame and set list of crop per season, downfall and temperature. The second part applied support vector matrices to prediction state for crop dataset. The SVM results calculate on regression format dataset. Declaration crop for downfall mean value: the primary part scan information frame and set list of crop per season, downfall and crop.

3. CONCLUSION

Agriculture is that the most important application space notably within the developing countries like Asian country. Use of knowledge technology in agriculture will be true of higher cognitive process and farmers will yield in higher approach. Data processing plays an important role for higher cognitive process on many problems associated with agriculture field. It discusses concerning the role of knowledge mining within the agriculture field and their connected work by many authors in context to agriculture domain. It conjointly discusses on totally different data processing applications in resolution the various agricultural issues. This paper integrates the work of assorted authors in one place thus it's helpful for researchers to induce info of current situation of knowledge mining techniques and applications in context to agriculture field.

REFERENCES

[1] Yethiraj N G , "Applying Data Mining Techniques In The Field Of Agriculture And Allied Sciences", International Journal of Business Intelligent ISSN: 2278-2400, Volume 01, Issue 02,December 2012.

[2] Jain A, Murty MN, Flynn PJ. Data clustering: a review. ACM Comput Surv. 1999; 31(3):264–323.

[3] Mucherino, A., Papajorgji, P., &Pardalos, P. (2009), "Data mining in agriculture" (Vol. 34), Springer.

[4] Han J, Kamber M. Data mining: concepts and techniques. Massachusetts: Morgan Kaufmann Publishers; 2001.

[5] Zaki, M J (1999), "Parallel and distributed association mining: A survey". IEEE concurrency, 7(4), 14-25.