Survey on Crop Yield Prediction Using Agriculture Dataset

Vaishnavi S[1], Jayapritha S[2],Dharani M[3], Gowtham S[4]
1 Assistant Professor, Department of CSE, SNS College of Technology
2, 3, 4 Final Year, Department of CSE, SNS College of Technology

ABSTRACT-- **Agriculture plays a major role in Indian economy and it is a main source of income. Data mining plays an important role in agriculture. The proper selection of crops for sowing is very important for good yield prediction. In olden days farmer's experience on a particular field and crops are considered for yield prediction. So there is a need to transfer the huge amount of data available in the lab and agricultural universities into information which helps the farmers predict the good yield. This can be made possible with the data mining techniques.**

Keywords— **Put your keywords here, keywords are separated by comma.**

I.INTRODUCTION

India is mainly an agriculture country. It is the most important occupation for most of the Indian families. Majority of research works are based on the mechanism to identify the crop growth and improve its yield. In agriculture, information technology plays a vital role. In recent years new constraints and methods are introduced by different emerging application domains. The various parameters such as seed type, crop variety and environmental parameters such as sunlight (temperature), soil (ph), water(ph) rainfall and humidity. In agriculture, consistent productivity of crop yield depends on the capability of soil to posses physical, chemical and biological activities. The soils in India are mostly deficient in nutrients like nitrogen, Phosphorous and Potassium. The cost of the food can be reduced by increasing the food production and at the same time it should give benefits to the farmers. This is the main aim of any technology in agriculture.

Data mining is the process of extracting information and transforming into an understandable structure for future use. It is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records(cluster analysis),unusual records(anomaly detection),and dependencies (association rule). It plays an important role in agriculture database for extracting knowledge and for updating information about crops. One of the main causes that leads to losses worth billions of dollars every year is Drought.

II.DATA MINING

Formal way of extracting knowledge from the data set is the process of mining. With the rough data, a valuable and hidden knowledge can be extracted. Data Mining is one of the processes of Knowledge discovery in databases (KDD).
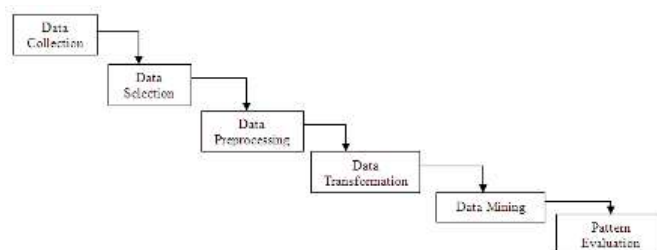
The steps KDD are:

Fig 1 Data mining Process

Steps are :

- Data Integration
- Data Selection
- Data Cleaning
- Data Transformation
- Data Mining
- Pattern Evaluation and Knowledge Presentation
- Use of Discovered Knowledge

A. DATA INTEGRATION

The datas are collected from different source and integrated .

B. DATA SELECTION

The data which is useful for our Data mining process is taken from the collected data.

C. DATA CLEANING

The collected data may contain wrong and irrelevant values. So, different techniques are applied to clean the data.

D. DATA TRANSFORMATION

Even the data which is cleaned are not suitable for mining process so it should be transformed into appropriate form by using the techniques like smoothing, aggregation, normalization etc.

E. DATA MINING

Now the transferred data is ready for Data mining process to discover the patterns by applying Data mining techniques like classification, clustering and association etc.

F. PATTERN EVALUATION AND KNOWLEDGE PRESENTATION

In this step visualization, transformation, removing redundant patterns from the generated patterns are carried out.

G. USE OF DISCOVERED KNOWLEDGE

The knowledge acquired is used to take better decisions.

III.TECHNIQUES IN DATA MINING

A. ASSOCIATION

Association is one of the techniques used in Data mining process to find the correlation between two or more items.

For example, the shopkeeper have an idea that the customer who buys bread have the chances of buying jam by tracking the purchasing details of the customer. Therefore the shopkeeper plans to keep the bread and jam nearby in the store.

B. CLASSIFICATION

In classification technique, the multiple attributes are already defined to identify the particular class. So that it is easy to choose the type of customer, item or an object.

For example , we can identify the model of the car by using the attributes like number of seats, car shape, wheels etc which is already defined.

C. CLUSTERING

Clustering is another type in Data mining techniques. It is same like classification techniques but the difference is that in clustering grouping was done based on their similarities. The attributes for grouping are not already defined.
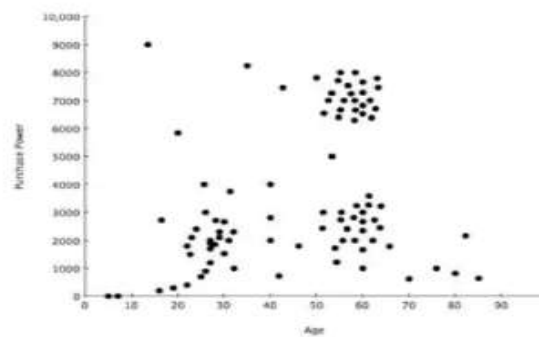

Fig 2 Example for Clustering

In example the two clusters are based on the people of age group between 20-30 can purchase for US$2,000 and another age group between 50-65 can purchase for US$7,000-8,000

D. DECISION TREES

Decision tree can be used for selection process. In which the top most node is the root node and the internal node contain the test on the attribute and the leaf node contains the class label. It is flow-chart like structure. Each answer leads to another question to identify the data.
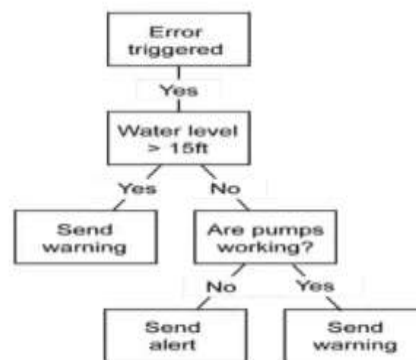

Fig 3 Decision tree

IV.METHODOLOGIES

Data mining plays an important role in agriculture field and it is the recent research area. The research article [1] *Pritam Bose* used the spiking neural network(SNN) for remote sensing spatiotemporal

analysis of image time series by using Moderate Imaging Resolution Spectroradiometer(MODIS) and Normalized Difference Vegetation Index(NDVI).

A. SPIKING NEURAL NETWORKS (SNNs)

It is the third generation of neural network models, increasing the level of realism in the neural simulation.

Moderate Imaging Resolution Spectroradiometer (MODIS)

It is an instrument that captures data in 36 spectral bands ranging in wavelength from 0.4µm to 14.4µm and at varying spatial resolutions. Together the instruments image the entire earth every 1 to 2 days.

Normalized Difference Vegetation Index (NDVI)

The simple graphical indicator used to analyze remote sensing measurements. Negative values of NDVI (Value approaching -1) correspond to water. Values close to zero (-0.1 to 0.1) are used to indicate that the area is of rock, sand, or snow. Lastly low positive values are used to represent the area which has shrub and grassland (approximately 0.2 to 0.4).while high values indicate temperate and tropical rainforests.

Experimental analysis shows the accuracy (95.64%) obtained for crop yield prediction using MODIS and NDVI was satisfactory. Every year NDVI images are collected for same duration by using Spectroradiometer. From the research article [2] they used data mining techniques to predict the category of soil data set. They used Naive Bayes and K-Nearest Neighbour methods to predict the crop yield.

B. NAÏVE BAYES CLASSIFIER

In machine learning Naïve Bayes classifiers are a family of simple probabilistic classifiers based on applying Baye's theorem with strong (naïve) independence assumptions between the features. It is well suited in complex real world situation.

C. K-Nearest Neighbor

K-Nearest Neighbor is shortly called as KNN algorithm. This is used for both classification and regression. The output of this method is based on whether it is used for classification or regression

technique. It is mostly used for classification techniques and it is easy for interpretation and takes low time for calculation. They applied the data mining techniques to classify the soil into low, medium and high category to predict the crop yield. The soil under medium category have good crop yielding capacity and the soil under high and very high category showed moderate yield capacity while the soil under low and very low category showed poor yielding capacity.

From the article [3] they integrated the crop growth model and remote sensing observation to investigate the effects of drought on crop yield. To understand the impact of drought on agriculture crop yield SMOS SM product at 25km is downscaled to 1km and assimilated into DSSAT crop growth model. Crop Yield was predicted by using the soil moisture and ocean salinity.

From the article [4] they considered soil to increase the crop yield. Its fertility rate was predicted by applying different classification algorithms.

D. J48

J48 is an open source java implementation of the C4.5 algorithm in the Weka data mining tool. C4.5 is an algorithm used to generate a decision tree. Its is an extension of Quinlan's earlier ID3 algorithm.

E. WEKA

It is a machine learning software written in Java, developed at the University of Waikato , New Zealand. It is a free software licensed under the GNU General Public License. Better result is obtained while using J48 algorithm. Fertilizers can be recommended to the farmers based on the decision tree form by J48 algorithm. J48 algorithm are implemented by using WEKA (Waikato Environment for Knowledge Analysis) tool.

From the research [5] they used bee hive modeling approach to analyze and classify the crop growth pattern and yield.

F. Bee Hive

It is based on the intelligent foraging behavior of honey bee swarm and it is an

optimization algorithm. It is used to solve many kinds of problems.

Types of bees are:

- Employed bees
- Onlooker bees
- Scout bees

EMPLOYED BEES

This bees search for food around the food source in their memory and they share the information about the food source to the onlooker bee.

ONLOOKER BEES

The onlooker bees found the food source with the help of employed bees. They select the food source which has high quality than the food with low quality. Here quality refers to fitness.

SCOUT BEES

These bees are transformed from the employed bees after finishing their food source and searching for another new food source. In this algorithm the number of employed bees is equal to the number of onlooker bees.

V.CONCLUSION

Crop yield prediction plays a significant process for a farmer to decide for agriculture. By using these techniques, Agriculture blooms to great extent. Data mining has a key role in predicting it. Various algorithms of mining can be applied to have a high impact of crop yield.

A. *REFERENCES*

[1] "Spiking Neural Network for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series",Pritam Bose, Nikola K. Kasabov, Lorenzo Bruzzone and Rreggio N. Hartono, *IEEE transactions on Geoscience and Remote sensing*, 2016

[2] " Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach",Monali Paul, Santosh K. Vishwakarma, Ashok Verma,*2015 International Conference on Computational Intelligence and Communication Networks*

[3] "Assimilation of SMOS Soil Moisture for Quantifying Drought Impacts on Crop Yield in Agricultural Regions", Subit Chakrabarti, TaraBongiovanni, Lincoln Zotarelli and Cimelio Bayer, *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 7, NO. 9, SEPTEMBER 2014*

[4] "Comparative Analysis of Classification Techniques on Soil Data to Predict Fertility Rate for Aurangabad District", Vrushali Bhuyar*, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 3, Issue 2, March – April 2014*

[5] "CRY – An improved Crop Yield Prediction model using Bee Hive Clustering Approach for Agricultural data sets",M. Gunasundari Ananthara, Dr. T. Arunkumar, Ms. R. Hemavathy, *Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, February 21-22*