Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 8, Issue. 3, March 2019, pg.161 – 165

A Survey on Predictive Analysis in Agricultural Soil Health Data to Predict the Best Fitting Crop

H.D.Gadade[1]; Riddhi Singh[2]; Vaishali Chaudhari[3]

[1,2,3]Government College of Engineering, Jalgaon, India

[1] gadade4u@gmail.com; [2] riddhis139@gmail.com; [3] vaishalic288@gmail.com

Abstract*— Agriculture hold an important sector in the Indian economy as it contributes around 18% of India's gross domestic product (GDP). India is an agricultural based country where more than 50% of the population depends on agricultural. Hence there is a need to provide farmers with the effective technology and knowledge to yield better crops based on the type of soil. Different types of soil are present in India. Different types of soil have different mineral contents and each crop require different mineral components for their better growth. Each soil has certain specific characteristic and is suitable to grow only certain number of crops. Hence a farmer should know about the type of soil he possesses so that he can cultivate better crops. In this paper we have described various effective algorithms and neural network techniques which have been used to classify the soil data based on the mineral contents and predict the best suitable crop for it.

***Keywords*— Data mining, Neural Network, Agriculture, Soil Data Analysis, Classification**

I. INTRODUCTION

With the advances in the technology, size of the data being generated is huge. We can use this data to obtain the patterns that are of interest in numerous fields. The agricultural field is a field where application of the data mining would be beneficial to the farmers. Process of data mining includes discovery of patterns from large data sets. The characteristics of soil in a particular region make it more suitable for some crops. But repeated cultivation of same crops leads to decline in the soil fertility as well as buildup of chemicals which may alter the soil pH. To counter this, the alternate cultivation of the crops can be an effective measure. We can use the data mining process to the datasets available in agricultural field to predict the crop which is suitable for that particular soil.

In this paper, the data mining process is applied to the dataset of the soil analysis which includes characteristics of soil like soil type and properties like percentages of nutrients in the soil. For this, we use the classification technique in which unknown samples are classified using training sets. Training sets are the sets of classified samples used to train classification technique how to perform its classification. The extraction of the data can be done by using various algorithms like naïve bayes algorithm, One-R, K-nearest neighbor algorithm

etc. Neural network can also be used to harvest the data from the datasets. A neural network is a network or circuit of neurons, or in a modern sense, an artificial neural network, composed of artificial neurons or nodes. An artificial neural network is a network of simple elements called artificial neurons, which receive input, change their internal state (activation) according to that input, and produce output depending on the input and activation. We have used the artificial neural network to predict the crop that is most suitable for a particular soil based on its characteristics and mineral contents. This will help the farmers to select the crop which will give them better yield and at the same time, will help in the restoration of the soil fertility. The good example of this is the plantation of the legume plants which can fix atmospheric nitrogen as rotational crop or inter-crop between regular crops.

II. LITERATURE SURVEY

Michael Goebel and Le Gruenwald [1] Knowledge discovery in databases is a rapidly growing field, whose development is driven by strong research interests as well as urgent practical, social, and economical needs. While the previous few years' knowledge discovery tools are being used mainly in the research environments, hence sophisticated software products are now rapidly emerging. In this paper, an overview to the common knowledge discovery tasks and approaches to solve these tasks are given. We have proposed a feature classification scheme that can be used to study knowledge and data mining software in an efficient way. This scheme is based on software's general characteristics database connectivity, and the data mining characteristics. We have applied various feature classification scheme in order to investigate 43 software products, which are either research prototypes or commercially available. In this way at the end we can state what features we consider as important for knowledge discovery software to possess in order to accommodate its users effectively, and handle issues that are either not addressed or insufficiently solved yet. An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

V Ramesh and K Ramar [2] The problem of the knowledge acquisition and efficient knowledge exploitation is very popular in agriculture area. One of the recognised methods for knowledge acquisition is the methods of classification from the existing agricultural databases. Weather and soil characteristics play an important role in agricultural decision making process. This research aimed to establish if meaningful relationship by assessing the various classification techniques of data mining and applying them to a soil science database. A large data set of soil database is extracted from the soil science and Agricultural Department, Kanchipuram and National Informatics Centre, Tamil Nadu. The applications of data mining techniques have never been conducted for Tamil Nadu soil data sets. In this research we compare the various classifiers and the outcomes of this research can improve the management and systems of soil uses throughout a large number of fields which include agriculture, horticulture, environmental and land use management.

D. Ashok Kumar, N. Kannathasan [3] Data mining has emerged as the major research domain in the recent decades in order to extract implicit and useful knowledge. This knowledge can be comprehended by humans easily. Initially, using statistical techniques this knowledge extraction was evaluated manually and computed. Subsequently, due to advancement in the technology semi automated data mining techniques emerged. Such advancement was also in the form of storage which led to an increase in the demands of analysis. In such case, semi-automated techniques have become inefficient. Therefore, in order to synthesis knowledge efficiently automated data mining techniques were introduced. In this paper a survey of the available literature on data mining and pattern recognition for soil data mining is presented. In Agricultural soil datasets data mining is a relatively novel research field. Efficient techniques have been developed and are being used for solving complex soil datasets using data mining.

Han, J. Et al [4] The objective of this research is to present a systematic analysis, in which we seek to identify the concepts of intelligent city, before the participation of the population with the intention of recognizing the nature of this intelligence. Therefore, the research associates concepts such as citizen participation and social web, smart cities, identifying publications between 2012 and the first quarter of 2018 and questioning what the intelligent city is and how this concept is able to (re)organize the learning processes of the territory from the dynamics of the contemporary city. The qualitative analysis of the documents revealed an innumerable of definitions and related terms such as: smart, intelligent, ubiquitous, digital, knowledge, sustainable, crowd sourcing, innovative; structured in three types of approaches: technological focus, focus on the human resources and focus on the citizen-related governance from the following domains: public and regulatory information policies, media convergence, geographic information system, crowd computing, infrastructure management, real-time data mining and extraction, smart cities education, and social monitoring and control. In spite of the access to a great amount of data, we can verify that the concept of intelligent city is referenced by a significant number of researches, but, in smaller number, works that present the models of construction of a collective intelligence for the city. From this perspective, we can identify the need to recognize the technological

education interventions for communication between the individuals and the city. Because we believe that only through the implementation and management of techno-Edu communication ecosystems will be able to promote a culture of participation.

Monali Paul et al [5] Yield prediction is very popular among farmers these days, which particularly contributes to the proper selection of crops for sowing. This has made the problem of predicting the yielding of crops an interesting challenge. Earlier yield prediction was performed by considering the information including farmer's experience on a particular field and crop. This work presents a system, techniques in order to predict the category of the analyzed soil datasets which uses data mining. The category, thus predicted will indicate the yielding of crops. The problem of predicting the crop yield is formalized where Naive Bayes and K-Nearest Neighbor methods are used as a classification rule.

N Heemageetha [6] India has 60.0% of its total area for agricultural purpose. Agricultural sector is the backbone for developing countries like India. The contribution of agricultural sector to the GDP is 17%. By improving the agricultural sector the GDP of the nation could also be improved. The digital era render its support to the agricultural sector in wide variety of ways. Data mining supports decision making process and prediction. Agriculture needs the decision support system in variety of ways such as type of crop to be cultivated. And prediction technique for rainfall prediction, weather prediction, market price prediction etc., there were many researches going on to support the agriculture using data mining. Analyzing the soil parameters provides a major contribution to the support of the farmers. This paper explores various proposed algorithms for analyzing soil using data mining techniques.

Yogesh Gandge and Sandhya [7] India is a country where agriculture and agriculture related industries are the major source of living for the people. Agriculture is a major source of economy of the country. It is also one of the country which suffer from major natural calamities like drought or flood which damages the crop. This leads to huge financial loss for the farmers thus leading to the suicide. Predicting the crop yield well in advance prior to its harvest can help the farmers and Government organizations to make appropriate planning like storing, selling, fixing minimum support price, importing/exporting etc. Predicting a crop well in advance requires a systematic study of huge data coming from various variables like soil quality ,pH ,EC,N,P,K etc. As Prediction of crop deals with large set of database thus making this prediction system a perfect candidate for application of data mining. Through data mining we extract the knowledge from the huge size of data. This paper presents the study about the various data mining techniques used for predicting the crop yield. The success of any crop yield prediction system heavily relies on how accurately the features have been extracted and how appropriately classifiers have been employed. This paper summarizes the results obtained by various algorithms which are being used by various authors for crop yield prediction, with their accuracy and recommendation.

Manasa Manjunatha and Parkavi A. [8] Agriculture is the widely practiced job in India which has the major share in the country's Gross Domestic Profit (GDP). Agriculture in India is being practiced as traditional job because of which agriculture is not being practiced as the technology driven or technology-oriented job. As a result, the farming practice in India is not producing good economic outcome. Data mining in agriculture is one such emerging trend which aims at improving the farming practice by considering the crop yield data. This crop yield data can be subjected to weather data and soil data in order to analyze the individual or combined effect of weather parameters and soil type on crop yield. This paper highlights some of the data analytical techniques used to analyze the crop yield based on which the current experiment is inspired.

Vibha L et al [9] -Predictive soil modelling using geostatistical methods is a research concept in modern soil science and soil geography for the last two decades. One of the major reasons for this lack of soil spatial data is that the conventional soil survey methods are relatively slow, expensive and qualitative. Spatial data sets covering large areas like digital geo morpho-graphical maps, geological, land use, and climate data are available and these geo-datasets contain information about soil formation and resulting hydrologic variables etc. which are needed to extract relevant soil information.

In this paper we present an efficient hybrid model that was achieved by first clustering the data and then classifying it, and using the spatial conceptual information extracted from the environmental variables. This paper assists in the assessment of the status of food production associated with the land degradation and estimate indicators of soil nutrient mining by the country and region. The findings and conclusions of this paper result from the monitoring of the nutrient mining of agricultural lands in a country which have a direct implication on policy development. We have proposed a framework in which the soil is classified into different types, then a future work could be to predict soil fertility, based on which you can decide upon the fertilizers and suitable crops which can then be cultivated with expertise.

Niketa Gandhi et al.[10] Rice crop production contributes to the food security of India, more than 40% to overall crop production. Its production is reliant on favorable climatic conditions. Variability from season to season is detrimental in farmer's income and livelihoods. By improving the ability of the farmers in order to predict the crop for productivity under different climatic scenarios, can thus assist farmers and other stakeholders

in making an important decisions in terms of and crop choice agronomy. This study thus aimed to use neural networks to predict rice production yield and investigate the factors which are affecting the rice crop yield for various districts of Maharashtra state in India. Data was sourced from publicly available Indian Government's records for 27 districts of Maharashtra state in India. The parameters which are considered for the present study were average temperature, maximum temperature and reference crop evapotranspiration, area, precipitation, minimum temperature, production and yield for the Kharif season in the month from June to November for the years 1998 to 2002. The dataset was processed using WEKA tool. A Multilayer Perceptron Neural Network was developed. To validate the data cross validation method was used. The results showed the accuracy of 97.5% with a sensitivity of 96.3 and specificity of 98.1. Further, the root mean squared error, relative absolute error mean absolute error, and root relative squared error were calculated for the present study. The study dataset was executed using Knowledge Flow of the WEKA tool. Using ROC curve the performance of the classifier is visually summarized.

III.METHODOLOGY

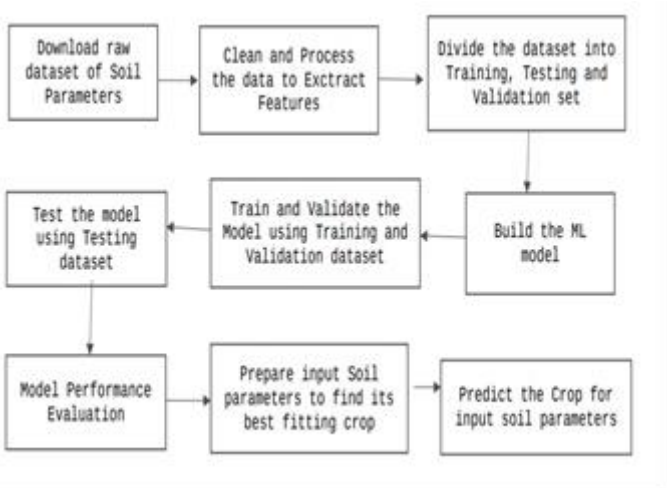

Fig: 3.1 Process Overview of Methodology

From the Fig. 3.1 we see that in this project, the input of program is the analysis of the soil sample provided. The analysis of the soil contains the constituents of the soil like the nitrogen content, phosphorous content, carbon content, etc.

The next step is to process the datasets formed by analysis of various soil samples and perform data cleaning on the datasets. In the cleaning of data, unavailable values are replaced with the mean of the values of that column. It also replaces negative and null values with average values. The different spellings of the entries in the soil characteristics dataset are merged. For eg. Red soil and red soil are merged in red_soil.

The processed data sets are then divided into training and testing datasets. 75% of the total data is included in the training dataset and 25% of the total data is included in the testing dataset. The size of the training and testing data can be changed as per our convenience.

The model of the project is created using numPy, keras and pandas. The model is used to train and test the data.

The training data set is used in multiple epochs to train the model. The 75% of the dataset is used for 1000 times to train the model so that the predicted answer is as close as possible.

The testing datasets are given as input to the model so as to check whether the training of the model is accurate or not. The testing datasets are 25% of the total data and can be used to check the result again and again.

After training and testing of data we can give the input of soil analysis for which the prediction is to be done.

Output of the program is the predicted crop suitable for the given soil sample analysis.

REFERENCES

[1] Michael Goebel and Le Gruenwald. A survey of data mining and knowledge discovery software tools. SIGKDD Explorations. Copyright @1999 ACM SIGKDD, June 1999.
[2] V Ramesh and K Ramar .Classifications of cultural land soils: A data mining approach. Agricultural Journal 6(3):82-86 (2011).
[3] Dr. D. Ashok Kumar*1, N. Kannathasan. A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining. IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 1, May 2011.
[4] Han, J., Pei, J. and Kamber, M. (2011) Data Mining Concepts and Techniques. Elsevier, New York.
[5] Monali Paul, Santosh K. Vishwakarma, Ashok Verma. Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining. 2015 International Conference on Computational Intelligence and Communication Networks
[6] N Heemageetha. A Survey on Applications of Data Mining Techniques to Analyze the soil for Agricultural Purpose. 978-9-3805-4421-2/16/$31.00 c 2016 IEEE.
[7] Yogesh Gandge and Sandhya. A Study on Various Data Mining Techniques for Crop Yield Prediction. 2017 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT).
[8] Manasa Manjunatha and Parkavi A. Estimation of Arecanut Yield in Various Climatic Zones of Karnataka using Data Mining Technique: A Survey. Proceeding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India
[9] Vibha L*, HarshaVardhan G M*, Prashanth S J*, P Deepa Shenoy*, Venugopal K R*, L M Patnaik**. A hybrid clustering and classification technique for soil data mining. lET-UKInternational Conference on Information andCommunication Technology in ElectricalSciences (ICTES 2007), Dr. MG.R. University, Chennai, Tamil Nadu, India. Dec. 20-22, 2007. pp. 1090-1095.
[10] Niketa Gandhi, Owaiz Petkar and Leisa J. Armstrong. Rice crop yield prediction using artificial neural networks. 2016 IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development (TIAR 2016).

