ISSN: 2321-7782 (Online)
e-ISJN: A4372-3114
Impact Factor: 7.327

Volume 5, Issue 11, November 2017

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Survey paper on Agriculture Yield Prediction Tool using Machine Learning

Niranjan G Hegde[1]
Student, Department of Computer Science
BMSCE
Bengaluru – India

Srisannidhi Mujumdar[2]
Student, Department of Computer Science
BMSCE
Bengaluru – India

Rajath Navada P R[3]
Student, Department of Computer Science
BMSCE
Bengaluru – India

Spoorti Shridhar Jambarmath[4]
Student, Department of Computer Science
BMSCE
Bengaluru – India

Prof. Madhavi R P[5]
Associate Professor, Department Of Computer Science
BMSCE
Bengaluru – India

Abstract: ***We intend to bring an enhancement in the field of precision agriculture by achieving better results in predicting crop yields as compared to the work already done in the field. With the use of machine learning techniques with proper optimizations and fine-tuned selection of the classifying algorithm, a system can be built that considers the data set comprising soil conditions, weather conditions of the past, builds a statistical model through learning and thus seeks to provide accurate and precise decision help with respect to the crops that can be grown profitably during the upcoming period of time. The output of this work would produce within the system, a set of rules (Knowledge base, which learns with more training from data sets) that helps the farmers pick most reliable crops to be grown for the present external factors, with least possible chances of losses.***

Keywords: Agriculture, Machine Learning, crop-prediction, Supervised Algorithms, Crop yield.

I. INTRODUCTION

There's always a significant risk factor to the farmers when deciding to grow a particular crop during a particular season, on a particular piece of land. Irrespective of the capital put in terms of soil nutrients, water and seed quality, the crop may fail bringing disastrous losses to the farmer and his family, eventually leading to more serious problems like debt and suicide. The exact reason though, may not be possible to be found out, often. In order to try and prevent such cases, there has been some significant research in the field of precision agriculture, by which attempts have been made to improve the crop production of farmers with the help of the data available from the past experiences of farming particular crops during particular seasons. But, the extents of their success ultimately depend on how well the real-time situations go with the kind of learning model the project uses and the factors that it considers in the data set. This field is a subject of a lot of research still, allowing umpteen scope to pick the combinations various strategies and factors apply them to bring out a model that suits the requirements the best.

II. Factors Influencing Crop Yield

There are several factors which influence the plant growth [1]. It includes mainly the soil nature that comprises of moisture [2], pH of soil [3], soil tilth [4] etc. And other factors such as irrigation salinity [5], rainfall and growing methods. These are the features that should be considered while predicting the crop yield.

III. Different Machine Learning Algorithms

K-nearest neighbours' algorithm: For classification and regression in pattern recognition we can use the k-n n algorithm as non-parametric based algorithm. 'k' closest training examples in the feature space can be used as inputs. Using this we can get output for both regression as well as classification. [6]

K-NN classifies the objects in to classes. The object is mapped to the class most common among its k nearest neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbour.[7] In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbours. [8]

Bootstrap aggregating, is a machine learning ensemble algorithm which helps other algorithms to improve the stability and accuracy. It also reduces deviations and helps to fine tune the curves. It is usually applied to decision tree methods, it can also be used with other methods. [9]

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. It takes the training data set to build an algorithm model that classifies the test set which includes the newer data , making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). In SVM model examples are plotted as points in space, in such a way that the examples of the different categories are distinguished by a clear gap that is as wide as possible. And test set examples are also categorized and mapped in to same space with the appropriate classification. [10]

A neural network is a highly complex interconnection of neurons(nodes or distributed network of processors), capable of 'learning' by examples and storing the information, which is further used to solve new decision or classification problems or situations.[11]Hence Neural networks can be used for machine learning and prediction in agriculture or any other well formatted data[12].

IV. Different Crop Prediction Methods

MATIS et al. (1985, 1989) proposed an alternative approach to forecast corn and cotton yield that used Markov Chain theory. This method overcomes some of the drawbacks of regression model. This method, being completely model free, does not require any assumptions about independent and dependent variables. [13]

In sensor based approach data is given as the input to the model which processes on it and gives output prediction on diseases and pest. Considering the raw data the variable assigned to the data null, over range, under range variables are deleted which is done during the Data Pre-processing stage. Then the Raw Data was categorized. E. Naive Bayes kernel algorithm is used to compare the patterns made of the crop data. A Sample Data Set of Crops consisted of data required for crops growth was given as training dataset to Naïve Bayes Kernel Model with input as raw data set of parameters soil samples and exterior temperature. Output from this model was pattern comparison of both data set. If the output pattern is consistent then there was no disease and growth of the crop is good and if it's inconsistent, diseases are predicted and necessary steps are taken. [14].

Crop Advisor: This an interactive website for finding out the influence of climate and crop production by using C4.algorithm.It gave the idea of how different climatic parameter influence the crop growth. Based on the C 4.5 algorithm, decision tree and decision rules have been developed. The selections were made based on the area under the chosen crop. The

information regarding the corresponding years climatic parameters such as rainfall, maximum and minimum temperature, wet day frequency where collected. [15]

Unsupervised Algorithms on Precision Agricultural Data: The method involved application of six machine learning algorithms to the Data set (Soya bean).The performance analysis of algorithms conducted on Weka Tools were done on the same dataset and constraints. The complexities of each algorithm were taken where DB Scan performed better, but K-means was more efficient for smaller dataset.[16]

The technique named CSM Is used to select the sequence of crops over the season. This method may improve the net yield of the crops. This method resolves the selection of crops for the particular season based on the prediction influenced by parameters such as weather, soil type etc. [17]

In one of the application data examined considered strawberry plants where the plant is controlling irrigation when the corresponding plant state is diagnosed. The prototype setup consisted of an array of 96 plants placed in a glass house. The plant and environmental parameters were: ETR (Electron Transport Rate by PAM meter), PAR (light by PARmeter), InfPAR (inflection PAR – a derived attributed by combining ETR and PAR), AmbC / PlantC (Ambient/Plant leaf temperature by thermistors), SM (Soil Moisture by probe EC- 10), and the learning goals Status (healthy/not healthy), Heat Stress (true/false), Drought Stress (true/false). Status was the characterization of the plant general stress status. Drought Stress is the characterization of the plant water stress status in relation to the soil moisture levels. Heat Stress can occur independently of water stress when the ambient environmental temperature gets very high and plant transpiration cannot maintain leaf cooling. Predictions were done by data collected on these. [18]

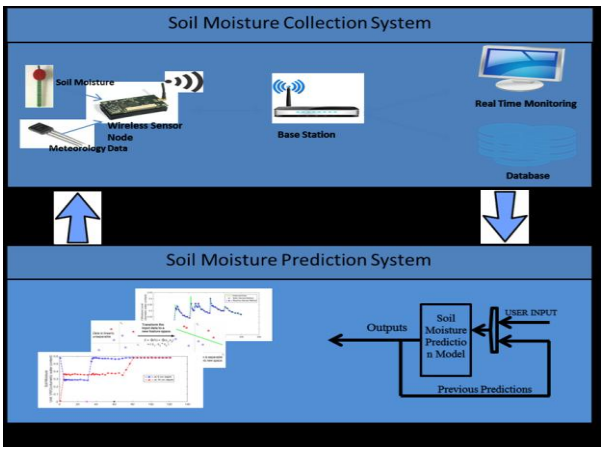

Fig1[19]

Fig. 1 shows an overview of the system, which was divided into two parts: collection and prediction. The design principle is to create a framework that allows users to easily configure the system to be site-specific.

In the collection system, a wireless sensor node is prototyped that implemented proposed framework for sensing soil moisture and other meteorological data. It offered two user-defined variables regulating the level of data granularity and sample intervals. The wireless sensor node can be used for applications such as in-field soil moisture collection and other kinds of remote site data collection, since it is specifically designed for applications that require a long lifetime. A prediction system is built on top of the machine learning models to predict soil moisture n days ahead. The models predict the soil moisture value based on meteorological parameters including temperature, humidity, wind speed, solar radiation, precipitation, and soil temperature together with previous days' soil moisture values. The sparse and well-studied machine learning techniques SVM and RVM are applied on the historical data to derive mathematical models. Designed from a Precision Agriculture perspective, the site specific model is able to incorporate data from other sources at the granularity of one day. The feature of taking user-provided data makes the system more robust by allowing the model to interact with human knowledge or reliable soil moisture

data from other sources at fine granularity. However, the soil and meteorological attributes collected from the hardware devices are the same attributes that are used for deriving prediction models. [19]

V. CONCLUSION

This paper presents survey for different yield prediction methods. Agriculture prediction would help the agricultural bodies in order to support farmers for making meaningful and profitable decisions regarding crop selection and methods of agriculture.

ACKNOWLEDGEMENT

The work, reported in this paper, is supported by the college through the Technical Education Quality Improvement Program [TEQIP-III] of the MHRD, Government of India.

References

1. Environmental Facts Effecting Plant Productivity, Matshal A Haferkamp, Fort Keogh Reasearch Symposium,Sep 1987
2. Plant growth and soil moisture relationships, J.FBierhuizen, Unesco/NS/AZ/476, Madrid Symposum Paper N0.37, Paris 1 September 1959
3. Kirste H. On the growth of plants in acid soils.(129-194)
4. Influence of Autumn Tilth on Soil Physical Conditions andTiming of Crop Establishment in Spring J. agric. Eng ng Res. (1984) 29, 265-213
5. Salinity and irrigation method affect crop yield and soil Quality In watermelon growing, Published online 20 February 2008 in Wiley Inter Science (www.interscience.wiley.com) DOI: 10.1002/ird.358
6. Altman, N. S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression". The American Statistician.46 (3): 175–185. doi:10.1080/00031305.1992.10475879.
7. D. Coomans; D.L. Massart (1982). "Alternative k-nearest neighbor rules in supervised pattern recognition: Part 1. K-Nearest neighbor classification by using alternative voting rules". Analytica Chimica Acta.136: 15–27. Doi: 10.1016/S0003-2670(01)95359-0.
8. A Detailed Introduction to K-Nearest Neighbor (KNN) AlgorithmMay 17, 2010 by Saravanan Thirumuruganathan
9. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. ClassificationAnd Regression Trees. Wadsworth.
10. V. Vapnik and A. Chervonenkis, Theory of Pattern Recognition, Nauka, Moscow, 1974.
11. Neural Networks A comprehensive foundation, Simon Haykin
12. Neural Network Applications In Agricultural Economics Jianhua Chen
13. Biom. J. 34 (1992) 4, 501-511 Akademie Verlag Probability Model for Crop Yield ForecastingR. C. JAIN and RANJANAAG UWALI.A.S.R.I., New Delhi
14. [doi 10.1109_ICACCS.2017.8014714] 20 International Conference on Advanced Computing and Communication Systems (ICACCS -2017), Jan. 06 – 07, 2017, Coimbatore, INDIA
15. [doi 10.1109_iccci.2014.6921718] Veenadhari, S. Misra, Bharat; Singh, CD -- [IEEE 2014 International Conference on Computer Communication and Informatics (ICCCI) - Coimbatore, India (2014.1.3-2014.)
16. [doi 10.1109_ICIIECS.2015.7193070] Mehta, Parth; Shah, Hetasha; Kori, Vineet; Vikani, Vivek; Shukla -- [IEEE 2015 International Conference on Innovations in Information, Embedded and Communication System
17. [doi 10.1109_ICSTM.2015.7225403] Kumar, Rakesh; Singh, M.P.; Kumar, Prabhat; Singh, J.P. -- [IEEE 2015 International Conference on Smart Technologies and Â Management Â forÂ Computing, Communication
18. [doi 10.1109_pci.2008.30] Dimitriadis, Savvas; Goumopoulos, Christos -- [IEEE 2008 12th Panhellenic Conference on Informatics - Samos, Greece (2008.08.28-2008.08.30)]
19. [doi 10.1109_smartcomp.2016.7501673] Hong, Zhihao; Kalbarczyk, Z.; Iyer, R. K.

