INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

A Survey on Efficient Crop Yield Prediction Using Random Forest Algorithm

Ajithkumar S[1], Dr.M.Somu[2], Dr.N. Saravanan[3], Dr.A.Rajivkannan[4]

PG Student, Dept. of C.S.E, K. S. R. College of Engineering (Autonomous), Tiruchengode, Tamilnadu, India[1]

Professors, Dept. of C.S.E, K. S. R. College of Engineering (Autonomous), Tiruchengode, Tamilnadu, India[2,3,4]

ABSTRACT: Rainfall Prediction is the application of science and technology to predict the state of the atmosphere. It is important to exactly determine the rainfall for effective use of water resources, crop productivity and pre planning of water structures. Using different data mining techniques it can predict rainfall. Data mining techniques are used to estimate the rainfall numerically. Agriculture has the largest contribution in the GDP of our country. But still the farmer's don't get worth price of the crops. There isn't any framework in location to suggest farmer what plants to grow. This prediction will help the farmers to choose appropriate crops for their farm according to the soil type, temperature, humidity, water level, spacing depth, soil PH, season, fertilizer and months. This prediction can be carried out using Random Forest classification machine learning algorithm

KEYWORDS: Temperature, Rainfall , Humidity , Water level, Soil ph, Spacing depth, Crop Yield Prediction, Random forest algorithm

I. INTRODUCTION

Analysis of time series data is one of the important aspects of modern research in the domain of knowledge discovery. Time series data is collected over a specific period of time such as hourly, daily, weekly, monthly, quarterly or yearly. Data mining techniques can use this data to predict upcoming situations in various domains such as climate change, education, and finance etc. These techniques can be used to extract hidden knowledge from time series data for future use. Weather forecasting is very beneficial but challenging task. Weather data consists of various atmospheric features such as wind speed, humidity, pressure and temperature etc.

Prediction can be helpful in many ways such as planning the water resources management, issuance of early flood warnings, managing the flight operations and limiting the transport & construction activities. Accurate rainfall prediction is more complex today due to climate variations. Researchers consistently have been working to predict rainfall with maximum accuracy by optimizing and integrating data mining techniques

II.RELATED WORK

1. Ponce-Guevara, K. L., Palacios-Echeverria, J. A., Maya-Olalla, E.,Dominguez Limaico, H. M., Suarez-Zambrano, L. E., Rosero-Montalvo,P.D., Alvarado-Perez, J. C.[1]This work shows the use of Big Data and Data Mining techniques on vegetable crops data from a greenhouse by implementing the first version of a software tool, so called GreenFarm-DM. Such a tool is aimed at analyzing the factors that influence the growth of the crops, and determine a predictive model of soil moisture. Within a greenhouse, the variables that affect crop growth are: relative humidity, soil moisture, ambient temperature, and levels of illumination and CO2.

2.Jheng, T.-Z., Li, T.-H., Lee, C.-P.[2]In this study, support vector regression (SVR) model is used as the main method to predict rice yield. We integrate the climate data and agricultural output dataset in Taiwan from 1995 to 2015. The experiment results show that the performance of hybrid SVR models are better than the traditional SVR model in terms of the root mean square error (RMSE) and the correlation coefficient (CC). The average RMSE and CC of hybrid SVR models are 60 and 0.996, respectively.

3.Manjunatha, M., Parkavi, A.[3]

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.

4.Shakoor, M. T., Rahman, K., Rayta, S. N., Chakrabarty, A. [4] Farmers usually plan the cultivation process based on their previous experiences. Due to the lack of précise knowledge about cultivation, they end up cultivatingundesirable crops. To help the farmers take decisions that can make their farming more efficient and profitable, the researchtries to establish an intelligent information prediction analysison farming in Bangladesh. However, this way of farming hereis still at the initial stage. The research suggests area based beneficial crop rank before the cultivation process.

5.Grajales, D. F. P., Mejia, F., Mosquera, G. J. A., Piedrahita, L. C.,Basurto,C.[5]Agriculture has an important tradition in Latin America and large areas have maintained stable conditions ford ecades with constant practices and yield. Planning processes in those areas have not suffered any change for years and some countries like Colombia have not modified its rural policies in the last three of four decades; on the other hand, natural and environmental conditions have changed dramatically in the last years as a consequence of climate change and climate variability modifying established practices for traditional crops. Since farmers have suffered several effects on its production process they have realized the importance of supporting future

6.Shah, P., Hiremath, D., Chaudhary, S.[6] Voluminous and variety of disparate information is generated and consumed in agriculture domain at a high local city. In agriculture, information is available in the form of weather and soil conditions reports, GPS mapping, water resources, fertilizer/ pesticide use, field characteristics, and commodity market conditions. Big data technology has a huge potential to refer these information and produce comprehensive insight via Geo-spatial processing, remote sensing, advance analytics algorithms, cloud resources, and advance storage systems.

7.Garg, A.,Garg,[7] Time-Series archetypes have been administered for prediction in variegated and disparate domains like academic enrollments, weather forecasting, rice production etc., where in an observation is contrived or excogitated based upon its past observation. The cogitation of fuzzy time series was first unveiled by Song & Chissom. This paper is the result of motivation from our previously published research work, wherein we delineate are gression based fuzzy time series approach, which deals with qualm, obfuscation,veracity, and spuriousness the various facets of the fuzzy contexture and furnishes a better and a more veracious result than the methods that are de facto and indubitable.

8. V. Sharmila, G. Tholkappia Arasu, P. Balamurugan[8] Non-class element based iterative clustering method was addressed. In this methodclasses are based on the weight calculation.

9. V.Vennila , A. Rajiv Kannan[9] Parallel linguistic fuzzy rule with canopy MapReduce (LFR-CM) framework was introduced. The framework uses canopy MapReduce function to classify big data for sharing information in the cloud.

10. P.Balamurugan, T.Ravichandran, V.Sharmila [10]Grade-Based Data Gathering (GBDG) algorithm to minimize the energy consumption of wireless sensor networks.

11. V.Vennila , A. Rajiv Kannan[11] Discritized Support Vector Classification and Prediction (DSV-CP) model was proposed to provide efficient Big Data computation and information sharing in Cloud computing environment.

12. V.Sharmila, P.Balamurugan, V.Vennila, S.Savitha [12] Data verification scheme to identify the malicious data packets.

13. P. Balamurugan, M. Shyamala Devi, V. Sharmila [13] Optimized methods was discussed for securing data (OMSD) which is trust based weights and also completely about the attacks and some methods for secured data transmission.

14. V.Vennila , A. Rajiv Kannan[14] Parallel Symmetric Matrix-based Predictive Bayes Classifier (PSM-PBC) model is developed for efficient Big Data computation and information sharing in Cloud environment.

15. P. Balamurugan, M. Shyamala Devi, V. Sharmila [15] Score-based data gathering algorithm was introduced that provides a significant solution to maximize the network lifetime as well as minimum delay per round of data gathering.

II. METHODOLOGY

A. *Existing System:*

The advancement in information storage is providing vast   amounts of data. A huge data set of crop database is extracted. The database contains measurements of soil data from various locations. In addition to the research establishes whether Soils are Classified Using various data mining techniques. Comparison was made between Naive bayes   classification and analyze the most effective technique.

B. *Proposed System:*

This system proposes an approach to analyse large data set. This proposal gives a introduction to application of analysis in the massive data analysis in the field of agronomy. Data about weather, irrigation, and yield from several other sources (e.g. Meteorological station and irrigation-plan records) For past few decades are collected and analyzed to produce an output. Which has the highest productivity of each grain in their respective geographical conditions simultaneously, the data about weather, soil condition, moisture content, due factor etc. Are recorded. From these records the random forest model are trained to evaluate the perfect crop for the current geographical conditions.

III. CONCLUSION AND FUTURE WORK

 Rainfall, which is an important factor for the prediction of crop yield is difficult to estimate precisely. Climate factors may change due to other remaining variables which may influence the prediction of rainfall. Also, the proposed work makes use of rain forest algorithm to estimate crop yield which works on a set range rather than discrete values, therefore, the error in predicted rainfall data does not cause problems as long as the difference between actual and estimated values is not drastic. Our model can successfully predict crop yield for a given year when the rainfall and temperature values for the previous years is known.

REFERENCES

[1] Ponce-Guevara, K. L., Palacios-Echeverria, J. A., Maya-Olalla, E.,Dominguez Limaico, H. M., Suarez-Zambrano, L. E., Rosero-Montalvo,P.D., Alvarado-Perez, J. C. (2017). GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data mining techniques(First trial). 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM).

[2] Jheng, T.-Z., Li, T.-H., Lee, C.-P. (2018). Using hybrid support vector regression to predict agricultural output. 2018 27th Wireless and OpticalCommunication Conference (WOCC).

[3] Manjunatha, M., Parkavi, A. (2018). Estimation of Arecanut Yield in Various Climatic Zones of Karnataka using Data Mining Technique: ASurvey. 2018 International Conference on Current Trends Towards Con- verging Technologies (ICCTCT).

[4] Shakoor, M. T., Rahman, K., Rayta, S. N., Chakrabarty, A. (2017). Agricultural production output prediction using Supervised Machine
Learning techniques.2017 1st International Conference on Next Generation Computing Applications (NextComp).

[5] Grajales, D. F. P., Mejia, F., Mosquera, G. J. A., Piedrahita, L. C.,Basurto,C. (2015). Crop-planning, making smarter agriculture with climate
data. 2015 Fourth International Conference on Agro-Geoinformatics (Agro-Geoinformatics).

[6] Shah, P., Hiremath, D., Chaudhary, S. (2017). Towards development of spark based agricultural information system including geo-spatial data.

[7] 2017 IEEE International Conference on Big Data (Big Data).Garg, A., Garg, B. (2017). A robust and novel regression based fuzzy time series algorithm for prediction of rice yield. 2017 InternationalConference on Intelligent Communication and Computational Techniques (ICCT).

[8] V. Sharmila, G. Tholkappia Arasu, P. Balamurugan Non-Class Element based Iterative Text Clustering Algorithm for Improved Clustering Accuracy using Semantic Ontology, Asian Journal of Research in Social Sciences and Humanities, 2016.

[9] V.Vennila & Dr. A. Rajiv Kannan Hybrid Parallel Linguistic Fuzzy Rules with Canopy MapReduce for Big Data Classification in Cloud, International Journal of Fuzzy Systems(IJFS), January 2019.

[10] P. Balamurugan, T.Ravichandran, V.Sharmila Grade and Energy based Data Gathering Protocols in Wireless Sensor Networks, Asian Journal of Research in Social Sciences and Humanities, 2016.

[11]V.Vennila & Dr. A. Rajiv Kannan Discretized Support Vector Prediction Classifier for Big Data Computation and Information Sharing in Cloud, Asian Journal of Research in Social Sciences and Humanities, 2017.

[12]V.Sharmila, P.Balamurugan, V.Vennila, S.Savitha Information Retrieval and Recommendation Framework Using Maximum Matched Pattern Based Topic Models, International Journal of Innovative Research in Engineering Science and Technology, 2016.

[13] P. Balamurugan, M. Shyamala Devi, V. Sharmila Detecting malicious nodes using data aggregation protocols in wireless sensor networks, International Journal of Engineering & Technology, 2018.

[14] V.Vennila & Dr. A. Rajiv Kannan Symmetric Matrix-based Predictive Classifier for Big Data computation and information sharing in Cloud, Elsevier Computers and Electrical Engineering, 2016.

[15] Balamurugan, M. Shyamala Devi, V. Sharmila An energy minimizing score based optimal data gathering in wireless sensor networks, International Journal of Engineering & Technology, 2018.

[16] M.Somu, S. Vignesh, D. Surya (2020), 'Water Distribution System'International Journal of Research and Advanced Development (IJRAD) ISSN: 2581-4451 Volume 4, Issue 1, pp. 57-61

[17] M Somu, P Akshaya, M Gowtham (2019), 'Data Mining Approach for Automatic Discovering Success Factors Relationship Statements in Full Text Articles' International Journal of Engineering Research & Technology (IJERT) Vol.7 Special Issue 01 pp.1-7

[18] [8]Somu,M and DineshKumar M (2016), 'Resource Constrained Secure Distributed DatabaseManagement Scheme for WSN' published in International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST) Vol. 2 Special Issues 10.

[19]Somu, M. and Vinoth, A(2013)'An Efficient approach to minimized data collection delay in wireless sensor networks 'IJCSET International Journal of Computer Science and Engineering Technology , Volume 5, Issue 1, April 2014 ISSN 2229-3345 pp.469-473.

[20] Somu M, Amudha K, Beni Steena T{2020), 'Enhancement of Image Quality by Noise Reduction using Modified Clustering Algorithm' Test Engineering & Management ISSN: 0193-4120,Volume 83, Page No. 1587 – 1591.