Application of Data Analytics in Agriculture Sector for Soil Health Analysis - Literature Review

Samihan Deshmukh
Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India
samihan25@gmail.com

Devesh Dhannawat
Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India
devesh.dhannawat@gmail.com

Mohit Dalvi
Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India
dalvimohit@gmail.com

Pratik Gawali
Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India
pratikgawali06@gmail.com

Prof. S. R. Vispute
Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune,India
visputesushma@gmail.com

Ms. Snehal Kekane
MS Computer Science,
University of North Carolina, Charlotte
snehal.kekane94@gmail.com

Abstract -

Soil is the most important part of agriculture. This paper compares different prediction models for factors which effects on soil health. Soil moisture, nitrogen, phosphorus, potassium, organic matter, heavy metals content is important for farmers to determine how much irrigation is required, which type of crops can be grown in such soil, which fertilizers to use for better yield from soil. High content of heavy metal can degrade the quality of soil. Such type of soil is also less useful for crops. This paper compares Prediction of heavy metals present in soil with SVM, RF, ELM. Among these three RF is found out to be most accurate and stable. SVM gives best accuracy for prediction of soil moisture and predicting the soil nutrients like N, P, K.

Keywords -

NPK values, Soil moisture, Soil heavy content, Comparative analysis, Support vector machine (SVM), Random forest (RF)

I. Introduction:

India is a country where primary occupation of most of the population is Agriculture. But there are no advance technologies being used in this field. As most of the population depends on agriculture there should be some techniques to maximize production of agriculture. Soil is one of the important element which affect directly on crop production. This paper is mainly concern on soil health based on many papers. Now days machine learning algorithms is technique to predict the unknown values. Data mining plays a important role in field of agriculture. This paper compares different machine learning approaches which are used for predicting the health of soil. Health of soil is dependent on many factors like soil moisture, soil heavy metal content, soil salt content, NPK values of soil, evaporation rate of water in soil, soil temperature, etc. This collective comparison of different paper will be useful to implement advance techniques in traditional farming.

There are different types of soil with varying characteristics, texture and features. Soil has a large effect of geographical, environmental and weather conditions. The concentration of minerals in soil can be detected with the help of pH of soil. Global radiation levels, soil salt content, changing atmospheric temperature, soil temperature, soil water evaporation rate contain affect the crop growth.

Agriculture needs decision providing system to indicate which type of crops should be cultivated.

II. Literature work

A. Technical survey:

Table 1 shows the Comparison table for different machine learning models used for predicting soil moisture according to paper "Spatial Prediction of Soil Moisture Content in Winter Wheat Based on Machine Learning Model" [1] by Hongmei Nie et al.

Table 1 : Comparison of soil moisture prediction

Model	Soil Depth (in cm)	Accuracy (in %)
SVM	0 to 20	92.899
	20 to 40	92.656
Random Forest	0 to 20	87.632
	20 to 40	87.842
Back-propagation Neural Network	0 to 20	80.570
	20 to 40	85.323

Table 2 shows the comparison table of related work of predicting soil fertility from paper "Machine Learning And Statistical Approaches Used In Estimating Parameters That Affect The Soil Fertility Status: A Survey" [2] by Sareena Rose et al.

Table 2 : Comparison of different soil parameter prediction

Objective	**Models**	**Best Model**
Predict CEC using silt, clay and sand contents, OC and pH	GEP, NF, NN, SVM	NF
Design PTF to estimate S-index USING SOC, BD, content of Nitrogen	ANN, LR	ANN
Predict SOC using nine continuous variables and two	MLR, CART, RF	RF

categorical		
Predict microbial dynamics using 3 inputs temperature, pH, incubation period	ANN, SVR, WM-FIS, SC-FIS	SC-FIS
Predict Soil Water Capacity using physical and chemical properties	GEP, NF, NN SVM, MARS, RF, ML	NF
Produce point PTF, PC PTF to determine SWRC using certain physical and chemical properties of the soil	MLR, ANN, SVR, kNN	ANN
Predict Phosphorus content using clay, sand, organic matter and pH	ANN, GA, FIS, ANFIS, PCR	ANN, GA

Many authors have used R-squared and Root Mean Square Error for comparing accuracy of their models.

R-squared is a statistical measure that represents the goodness of fit of a regression model. The ideal value for r-square is 1. The closer the value of r-square to 1, the better is the model fitted.

Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are. In other words, it tells you how concentrated the data is around the line of best fit. Root mean square error is commonly used in climatology, forecasting, and regression analysis to verify experimental results. The formula is:

$$RMSE = \sqrt{\overline{(f-o)^2}}$$

Where:

- f = forecasts (expected values or unknown results),
- o = observed values (known results).

Table 3 shows the comparison of predicting heavy metals in soil with different machine learning algorithms from "Predicting Soil Heavy Metal Based On Random Forest Model" [3] by Weibo Ma et al. Table 3 (a) is for metal cd (Cadmium), Table 3 (b) is for metal as (Arsenic), Table 3 (c) is for metal pb (Lead).

Table 3 (a) : Cd prediction comparison

Model	Evaluation Indicator	Value
PLS	R^2	0.5362
	RMSE	20.9907
SVM	R^2	0.6025
	RMSE	19.4837
RF	R^2	0.6486
	RMSE	15.5256
ELM	R^2	0.7085

	RMSE	16.6493

Table 3 (b) : As prediction comparison

Model	Evaluation Indicator	Value
PLS	R^2	0.9431
	RMSE	1.3683
SVM	R^2	0.9720
	RMSE	0.8760
RF	R^2	0.9912
	RMSE	0.5327
ELM	R^2	0.9653
	RMSE	0.7037

Table 3 (c) : Pb prediction comparison

Model	Evaluation Indicator	Value
PLS	R^2	0.9071
	RMSE	2.6327
SVM	R^2	0.9622
	RMSE	1.5919
RF	R^2	0.9756
	RMSE	1.1694
ELM	R^2	0.9652
	RMSE	1.6501

Table 4 shows comparison of soil classification from "Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series" [4] by Sk Al Zaminur Rahman et al.

Table 4 : Comparison of soil classification algorithm

Model	Accuracy (%)
Gaussian SVM	94.95
Weighted k-NN	92.93
Bagged trees	90.91

Table 5 shows performance of two phase fuzzy based prediction of N, P, K values from "Two phase fuzzy based prediction model to predict Soil nutrient" [5] by Ravinder Singh et al.

Table 5 : Comparison of N, P, K prediction

Element	RMSE
N	0.32
P	0.34
K	0.41

B. Non - Technical Survey:

1) "Soil Moisture Prediction Using Machine Learning" [6] by Shikha Prakash et al for predicting soil moisture. They have used machine learning algorithm like multiple linear regression, support vector regression and recurrent neural networks for prediction of soil moisture for 1 day, 2 days and 7 days ahead. The performance was compared with R^2 and RMSE. The comparison shows multiple linear regression is better in providing RMSE and R^2 of 0.14 and 0.975 for day1 , 0.353 and 0.939 for day 2, 1.59 and 0.786 for day 7 ahead.

2) "The prediction model for soil water evaporation based on BP neural network" [7] LiliMa et al for predicting soil water evaporation rate. They have used BP neural network for prediction which gives high accuracy and high stability. Average temperature, relative humidity,

net radiation these factors are considered for finding the output.

3) “Soil Fertility Grading With Bayesian Network Transfer Learning” [8] by Hai-yang Jia et al describes methods for grading soil into different types so that suitable crops can be grown in it. They have used Bayesian Network based transfer learning algorithm. Algorithm includes structural learning parameter learning, considering both similarities between learning task and the geographical position of land square. Empirical experiment results show a significant improvement in terms of structure and parameters when transfer knowledge between similar soil fertility grading task.

4) “Predicting Soil Heavy Metal Based On Random Forest Model” [3] by Weibo Ma et al predicts heavy metal contain of soil. They have used three machine learning algorithms SVM, RF, ELM and compared with PLS method. Here ELM and RF performs better than SVM. The concentration of metals will affect the prediction of ELM. Stability of RF is best among these three models. RF algorithm has higher accuracy for inversion of soil heavy metal research.

III. Predictive Model Flow

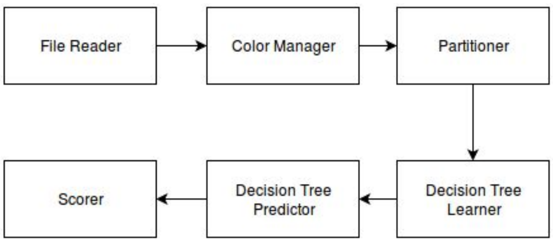

Figure 1 : Predictive Model Flow for Knime Tool

IV. Conclusion

In This literature survey for soil health analysis we compared prediction of different soil elements with different machine learning algorithms. This survey will be very useful for those who are building products related to soil health analysis and prediction. In the we came to conclude that for most of the cases support vector machine (SVM) performs better than other algorithms. Only for heavy metal prediction Random Forest (RF) performs better than SVM. This implies SVM should be preferred over other machine learning algorithms when dealing with soil health analysis and for prediction of heavy metals in soil Random Forest should be used. This paper will be responsible for increasing the accuracy of existing or upcoming models related to soil health prediction.

V. References

[1] H. Nie, L. Yang, X. Li, L. Ren, J. Xu and Y. Feng, "Spatial Prediction of Soil Moisture Content in Winter Wheat Based on Machine Learning Model," *2018 26th International Conference on Geoinformatics*, Kunming, 2018, pp. 1-6

[2] S. Rose, S. Nickolas and S. Sangeetha, "Machine Learning and Statistical Approaches used in Estimating Parameters that Affect the Soil Fertility Status: A Survey," *2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT)*, Bangalore, India, 2018, pp. 381-385

[3] W. Ma, K. Tan and P. Du, "Predicting soil heavy metal based on Random Forest model," *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Beijing, 2016, pp. 4331-4334

[4] Al Zaminur Rahman, Sk & Chandra Mitra, Kaushik & Islam, S.M.. (2018). Soil Classification Using Machine Learning Methods and Crop Suggestion Based on Soil Series. 1-4

[5] Singh, Ravinder & Kaur, Rubleen & Singla, Anshu. (2018). Two phase fuzzy based prediction model to predict Soil nutrient. 80-85

[6] Prakash, Shikha & Sharma, Animesh & Sahu, Sitanshu. (2018). Soil Moisture Prediction Using Machine Learning. 1-6

[7] L. Ma, Y. Wu, J. Ji and C. He, "The Prediction Model for Soil Water Evaporation Based on BP Neural Network," *2011 International Conference on Computer Distributed Control and Intelligent Environmental Monitoring*, Changsha, 2011, pp. 641-645

[8] H. Jia, J. Chen, H. Yu and D. Liu, "Soil fertility grading with Bayesian Network transfer learning," *2010 International Conference on Machine Learning and Cybernetics*, Qingdao, 2010, pp. 1159-1163

[9] S. R. Vispute and M. A. Potey, "Automatic text categorization of marathi documents using clustering technique," *2013 15th International Conference on Advanced Computing Technologies (ICACT)*, Rajampet, 2013, pp. 1-5

[10] E. Claverie, J. Lecoeur, V. Letort and P. Cournède, "Modeling soil temperature to predict emergence," *2016 IEEE International Conference on Functional-Structural Plant Growth Modeling, Simulation, Visualization and Applications (FSPMA)*, Qingdao, 2016, pp. 28-37

[11] Y. Ge, Z. Liu and B. Wang, "Soil Water Prediction of Moving Dune Based on BP Neural Network Model in Northwest Liaoning Sandy Land," *2010 Second International Conference on Computer Modeling and Simulation*, Sanya, Hainan, 2010, pp. 215-219

[12] P. A. Harlianto, T. B. Adji and N. A. Setiawan, "Comparison of machine learning algorithms for soil type classification," *2017 3rd International Conference on Science and Technology - Computer (ICST)*, Yogyakarta, 2017, pp. 7-10

[13] M. Paul, S. K. Vishwakarma and A. Verma, "Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach," *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, Jabalpur, 2015, pp. 766-771

[14] S. Peng, T. Li and F. Wang, "Time Series Prediction Model of Soil Moisture Based on Wavelet De-Noising," *2009 International Conference on Management and Service Science*, Wuhan, 2009, pp. 1-4

[15] E. Santi *et al.*, "Estimating Soil Moisture from C and X Band Sar Using Machine Learning Algorithms and Compact Polarimetry," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, 2018, pp. 1426-1429

[16] D. Smith and W. Peng, "Machine learning approaches for soil classification in a multi-agent deficit irrigation control system," *2009 IEEE International Conference on Industrial Technology*, Gippsland, VIC, 2009, pp. 1-6

[17] M. Veres, G. Lacey and G. W. Taylor, "Deep Learning Architectures for Soil Property Prediction," *2015 12th Conference on Computer and Robot Vision*, Halifax, NS, 2015, pp. 8-15

[18] Y. Liu, X. Pan, R. Shi, Y. Li, C. Wang and Z. Li, "Predicting Soil Salt Content Over Partially Vegetated Surfaces Using Non-Negative Matrix Factorization," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 11, pp. 5305-5316, Nov. 2015

[19] Yong He, Hai-Yan Song, Pereira and Gomez, "Study on spectral feature characterizing and soil nitrogen content prediction," *2005 International Conference on Machine Learning and Cybernetics*, Guangzhou, China, 2005, pp. 5424-5428 Vol. 9

[20] Yue Liu, Long Mei and Su Ki Ooi, "Prediction of soil moisture based on Extreme Learning Machine for an apple orchard," *2014*

IEEE 3rd International Conference on Cloud Computing and Intelligence Systems, Shenzhen, 2014, pp. 400-404

[21] M. Zhang, S. Wang, S. Li, J. Yi and P. Fu, "Prediction and map-making of soil organic matter of soil profile based on imaging spectroscopy: A case in Hubei China," *2011 19th International Conference on Geoinformatics*, Shanghai, 2011, pp. 1-5

[22] Mr. Harshad Bhujbal et al, "An Application For E-Marketing Of Agricultural Commodities And Analysis Of Marketable Surplus" International Journal of Advanced Research in Computer and Communication Engineering, Vol. 7, Issue 4, April 2018

[23] Prachi S., Kirti P., Pooja N., Alisha S., Sushma V. (2018) Automatic Integration and Clustering of Marathi Documents in Different Formats for Effective Information Retrieval. In: Tiwari B., Tiwari V., Das K., Mishra D., Bansal J. (eds) Proceedings of International Conference on Recent Advancement on Computer and Communication. Lecture Notes in Networks and Systems, vol 34. Springer, Singapore

[24] Vinit Savant et al "A Survey On Farmer's Need and Feedback Analysis System" International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 11, November 2015

[25] R. Vispute, Sushma & Patil, Shrikant & Sangale, Sagar & Padwal, Akshay & Ukarde, Avinash. (2015). Parallel Processing System for Marathi Content Generation. Proceedings - 1st International Conference on Computing, Communication, Control and Automation, ICCUBEA 2015. 575-579. 10.1109/ICCUBEA.2015.118.

[26] Sirsat, Manisha & Cernadas, E & Fernandez-Delgado, M & Barro, S. (2018). Automatic prediction of village-wise soil fertility for several nutrients in India using a wide range of regression methods. Computers and Electronics in Agriculture. 154.120-133.10.1016/j.compag.2018.08.003.

[27] Vispute, S. R., Patil, S., Sangale, S., Padwal, A., & Ukarde, A. (2015). Parallel Processing System for Marathi Content Generation. 2015 International Conference on Computing Communication Control and Automation