Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7765922

ANALYSIS AND PREDICTION FOR MICRONUTRIENTS PERFORMANCE IN PADDY USING DATA MINING AND MACHINE LEARNING APPROACHES

S. Dhanavel[1] and A. Murugan[2]

[1]Research Scholar, Department of Computer and Information Science, Annamalai University, Annamalainagar – 608 002, Tamil Nadu, India
Email: dhanavel2008@gmail.com

[2]Assistant Professor, Department of Computer Science, Thiru Kolanjiappar Govt. Arts College, Virudhachalam, Tamil Nadu, India
Email: drmuruganapcs@gmail.com

Abstract

In this study, data mining and statistical approaches were established to estimate the agriculture growth using soil micro and macro nutrient level. Agricultural development covers different parameters like weather conditions and soil nutrient level. In this paper, consider four essential plant nutrient elements defined as micronutrients namely, zinc (Zn), iron (Fe), copper (Cu) and manganese (Mn). To archive this objective, analysis of paddy yield and level of soil micronutrients using data mining and machine learning approaches with stochastic gaussian process, linear regression model, and sequential minimal optimization algorithm. Numerical illustrations also provide to prove the results and discussions using different nutrients parameters and various machine learning classifiers with its accuracy parameters namely R2 score, Mean Absolute Error (MAE), Root Mean Squar Error (RMSE), Relative Absolute Error (RAE) and Root Relative Squared Error (RRSE).

Index Terms: Data Mining, Machine Learning, Gaussian Process and Performance Metrics.

1. Introduction

In India, the field of agriculture for most important for increasing the Indian economy. The current situations the field of agriculture face different problems like weather conditions, labour demands and soil nutrient level. Indian soils have become deficient not only in major plant nutrients like nitrogen, phosphorus and in some cases, potash but also in secondary nutrients, like sulphur, calcium, and magnesium. Micronutrients such as zinc, boron and to a limited extent iron, manganese, copper and molybdenum have also been reported to be deficient.

Deficiency of micronutrients during the last three decades has grown in both, magnitude, and extent because of increased use of high analysis fertilizers, use of high yielding crop varieties and increase in cropping intensity. This has become a major constraint to production and productivity of paddy, sugarcane, and wheat. Thus, there is an urgent need for correction of individual nutrient deficiency and for arresting its further spread.

In this research, collect different data regarding agricultural related primary and secondary data for various resources like department of economics and statistics, department of agriculture and other relevant departments for state and central government. Data collection faces different stages in the area of data mining and also finding different hidden information using various statistical analysis. Finally, based on numerical illustrations how to increase the paddy growth using micronutrients. There are four essential micronutrients in plants. Some

nutrients control the permeability of a cell membrane and some others control the osmotic pressure, buffer action, etc. zinc (Zn), iron (Fe), copper (Cu) and manganese (Mn) are some of the important micronutrients in plants. Lack of any of these nutrients affects growth and development. Many of these deficiencies are preventable through increasing the level of nutrition.

Machine Learning is a technology that provides systems with the ability to automatically learn and improve from experience by repeated training. It includes a set of well-defined models that collect specific data and apply specific algorithms to achieve desired results. Machine learning techniques have been applied to agriculture domain in order to improve the productivity and quality of the crops grown. The algorithms in Machine Learning are used to determine for a particular crop under which conditions the best yield would be produced.

Crop Yield prediction depends on various factors like the soil composition, type, climate, regional geography and disease or pests. Soil is a very important factor affecting plant growth. It consists of solids, liquids and gases and contains living organisms. All these elements provide their physical and chemical properties. To maintain fertility, to achieve better yield, and to protect the environment; it is necessary to nurture the soil properly. The analysis of soil nutrients is very useful for the farmers in determining the type of yield to be grown in a particular soil condition. Good soil fertility management requires careful identification of the limits of current nutritional deficiencies and monitoring of changes in soil fertility to predict their shortage. These gaps must be mitigated by sound and best practices in terms of nutrients, water, plants and energy for soil management, in order to maintain food production at a reasonable level to ensure high productivity at the same time. Therefore, managing soil fertility at optimal levels is one of the key factors for achieving high and sustainable productivity (Smriti 2015).

Rapeseed-Mustard is the second most important oil seed crop. Productivity of rapeseed mustard in J&K remains unstable from past few years (Rakesh 2018). The current research work makes use of supervised machine learning algorithms to predict mustard crop yield for different districts of Jammu region. In this work, the authors implemented different ML techniques for crop yield prediction to find out the most accurate techniques for crop yield prediction under study. This research paper is organized as follows: Section 2 represents Literature Review, Section 3 explains Materials and Methodology, Section 4 shows Results and Discussion, and Section 5 describes Conclusion and Future Scope.

2. Literature Review

In recent years, machine learning techniques have been applied in agriculture domain by various researchers. A review on the implementation of different ML techniques in the field of crop yield prediction from soil analysis from past few years is presented as under. Bhuyar (2014) proposed an approach where different classification algorithms such as J48, Naïve Bayes, and Random forest algorithm were applied to soil data set to predict its fertility. J48 algorithm gave better result with an accuracy of 98.17% than other algorithms. After two years, Rajeshwari and Arunesh (2016) performed a comparative analysis of ML algorithms i.e. Naive Bayes, JRIP and J48 for prediction of soil types. The experiments were performed on soil data consisting of 110 samples using data analytics tool R. The experimental results predicted that

JRIP algorithm performed better as it gave highest accuracy of 98.18% with kappa statistic of approximate 1.0.

In the same year, Sujata (2016) proposed a model to estimate the crop yield in order to improve the value and gain of farming area using data mining techniques. Awasthi and Bansal (2017) performed comparative study on two data mining techniques namely Artificial Neural Network and Support Vector Machine with the help of data analytics tool R. ANN was implemented with 7 hidden nodes and this model trained for 73073 steps. It predicted accuracy of 55% with root mean square error is of 15. SVM implemented with Radial basis kernel and it achieved much better results with 74% of accuracy.

In the same year, Singh et al. (2017) implemented different machine learning techniques in order to predict the category of the rice crop yield based on macro-nutrients and micro- nutrients status in dataset. ANN, Naïve Bayes and KNN are applied on soil data. Decision Tree Classifier and Naïve Bayes Classifier are found to be better models for classifying the soils into categories and in the prediction of yield on the basis of Nutrient status in the soil. Supriya (2017) presented a system based on data mining techniques in order to predict the category of the analyzed soil datasets (yield of crop). Naive Bayes and K-Nearest Neighbor methods are used. System architecture is also developed and tested using data mining technology.

Verma et al. (2018) also proposed an approach for wheat crop prediction using fuzzy-c means clustering and neural network. In the same year, Priya et al. (2018) implemented Random Forest technique for crop yield prediction in Tamil Nadu state. The dataset for the study includes various parameters such as rainfall, temperature, crop production, etc. and experiments were performed using R Studio. Next year, Renuka and Terdal (2019) applied machine learning techniques namely KNN, SVM and Decision Tree for yield prediction of sugarcane crop. The study was carried out in Python platform. Decision Tree predicted highest accuracy of 99% with less mean square error.

Jayalakshmi and Devi (2019) applied ML techniques for predicting soil fertility. The authors implemented C5.0, Random Forest and K-Nearest Neighbor for crop production with high accuracy and efficiency by generating a model which predicts whether soil is “Ideal” or “Not Ideal” for crop production based on soil parameter. C5.0 predicted highest performance with an accuracy of 96%.

3. Materials and Methodology

The main objective of the current research work is to predict crop yield from soil analysis using machine learning techniques. To attain the objective of the current research, experiments have been carried out using WEKA and Python platform. After gaining insight into problem domain, discussion with farmers and soil chemists and reviewing literature; research problem has been framed out. For current research problem, real data has been collected Rokiyaraj et al. (2016) and Department of Agriculture. Data has been collected from The study area covers 22 revenue villages in coastal rice growing agriculture land area of Tharangambadi taluk in Nagapattinam District in Tamil Nadu. This dataset consists of 22 instances with 4 input parameters representing soil micronutrient status of Nagapattinam District in Tamilnadu region. The parameters of the dataset are zinc (Zn), iron (Fe), copper (Cu) and manganese (Mn) representing soil micronutrients. The instances of the dataset are presented in table 1.

Table 1. Paddy Area Micronutrients Level in Tharangambadi taluk in Nagapattinam District in Tamilnadu

Village Name	Zn (mg kg)	Fe (mg kg)	Cu (mg kg)	Mn (mg kg)
Koodalore	1.68	14	1.8	2.75
Eravancheri	1.7	19.5	1.72	2.8
Nalladai	1.68	18	1.67	2.85
Marudhampallam	1.7	9.65	1.81	2.89
Mamakudi	1.68	20.45	1.43	2.7
Kidangal	1.58	22.4	1.84	2.9
Memathur	1.69	21	1.7	3.1
Mathur	1.5	19	1.8	2.7
Thirukadaiyur	1.69	18	1.79	3
Pillaiperumanallore	1.4	19	1.9	2.9
Manikkapangu	1.05	18.5	2.03	2.01
Erukattacheri	1.56	18	2.6	1.9
Ilupur	0.6	9.5	1.74	2.9
Eduthukatti	1.3	6.5	1.5	2.6
Thirukklacheri	1	6	1.59	2.5
Thiruvidaikazhi	1.02	6.4	1.05	2.9
Thillaiyadi	1.3	7.8	1.6	3
Kattucheri	0.9	9	1.68	2.53
Poraiyar	0.82	7	1.9	3.3
Chandrapadi	0.9	3.5	1.69	2.8
Melaperumpallam	1	7	1.53	2.9
Keelaperumpallam	1.11	6.9	1.52	2.89
Radhanallur	0.778	4.042	0.758	1.89
Arkkadu	0.77	3.87	0.77	1.86
Komal	0.646	4.35	0.736	1.78
Panithalamedu	0.644	4.39	0.744	1.78
Kizhamaruthathanallur	0.632	4.39	0.74	1.74
Anathadavapuram	0.628	4.02	0.784	1.74
Muvallur	0.65	4.06	0.92	1.76
Vanathirajapuram	0.85	4.75	0.898	2.35
Thirumanacherry	0.81	3.866	0.88	3.04
Mudikondanallur	0.784	3.694	0.866	3.288

After data collection, pre-processing has been performed to transform nominal values into numerical form, to input missing data and to detect outliers present in the data using statistical technique. The overall methodology for the proposed work is presented in figure 1. The experiments have been carried out in different phases. In the first phase completed using data pre-processing. The second phase deals with implementation of Gaussian Process respectively for training nutrients classifier. In each phase, data (with selected features) has been divided into two sets: training set ration of 70% and test set ration of 30% samples. Each trained model has been validated with test data in order to assess its performance. The performance of all the trained models has been measured using performance measures namely MAE, RMSE, RAE, and RRSE. On the basis of the experimental results, comparative analysis of all the trained models has been carried out to reveal the most accurate technique for nutrients level prediction.

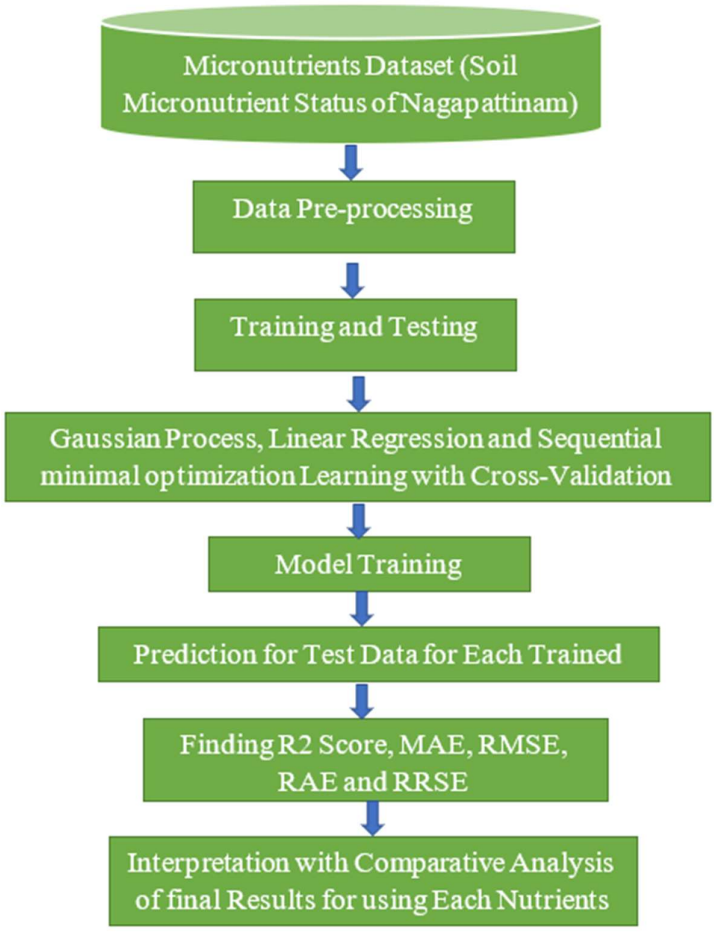

Fig. 1. Proposed System Model

Wang (2020) and Rasmussen (2006) explain the concepts of gaussian process is a machine learning technique. Use it to do regression, classification, among many other things. Being a Bayesian method, Gaussian Process makes predictions with uncertainty. Gaussian Process can be explained in a couple of formulas. The Gaussian processes model is a probabilistic supervised machine learning framework that has been widely used for regression and classification tasks. A Gaussian processes regression (GPR) model can make predictions incorporating prior knowledge (kernels) and provide uncertainty measures over predictions.

3.1 Gaussian processes

The Gaussian processes model is a supervised learning method developed by computer science and statistics communities. Researchers with engineering backgrounds often find it difficult to gain a clear understanding of it. To understand GPR, even only the basics need to have knowledge of multivariate normal distribution, kernels, non-parametric model, and joint and conditional probability. A random variable X is Gaussian or normally distributed with mean μ and variance $\sigma 2$ if its probability density function (PDF) is Murphy (2012).

$$P_X(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \ldots (1)$$

Here X represents random variables and x is the real argument. The normal distribution of X is usually represented by $P_X(x) \sim N(\mu, \sigma^2)$.

3.2 Linear regression

In the simplest words explain by analyticsvidhya, the Linear Regression is the supervised Machine Learning model in which the model finds the best fit linear line between the independent and dependent variable i.e. it finds the linear relationship between the dependent and independent variable. Linear Regression is of two types: Simple and Multiple. Simple Linear Regression is where only one independent variable is present and the model has to find the linear relationship of it with the dependent variable. Whereas, In Multiple Linear Regression there are more than one independent variables for the model to find the relationship.

$$Y = b_0 + b_1 X$$

Equation of Simple Linear Regression, where b_0 is the intercept, b_1 is coefficient or slope, X is the independent variable and y is the dependent variable. A Linear Regression model's main aim is to find the best fit linear line and the optimal values of intercept and coefficients such that the error is minimized. Error is the difference between the actual value and Predicted value and the goal is to reduce this difference. Let's understand this with the help of a diagram.

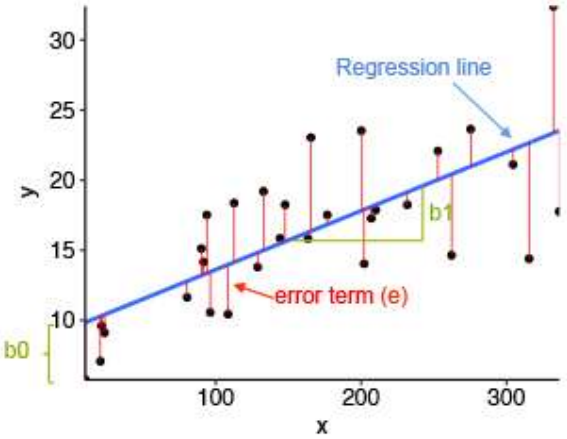

Image Source: Statistical tools for high-throughput data analysis

3.3 Sequential Minimal optimization

Jithin (2020) explain Sequential Minimal optimization (SMO) is an iterative algorithm for solving the Quadratic Programming (QP.) problem that arises during the training of Support Vector Machines (SVM). SMO is very fast and can quickly solve the SVM QP without using

any QP optimization steps at all. Consider a binary classification with a dataset $(x_1,y_1),\ldots..,(x_n,y_n)$ where xi is the input vector and $y_i \in \{-1,+1\}$ which is a binary label corresponding to each xi. An optimal margin SVM is found by solving a QP. problem that is expressed in the dual form as follows:

$$\max_{a} = \frac{1}{2}\sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \, K(x_i, x_j)\alpha_i\alpha_j$$

Subject to

$$0 \le \alpha_i \le C \;\; for \;\; i = 1, 2, \ldots, n$$

$$\sum_{i=1}^{n} y_i \alpha_i = 0$$

Here C is SVM hyperparameter that controls the tradeoff between maximum margin and loss and $K(x_i, x_j)$ is the Kernel Function. αi is Lagrange Multipliers. SMO is an iterative algorithm, and, in each step, it chooses two Lagrange Multipliers to jointly optimize and then finds the optimal values for these multipliers and updates the SVM to reflect the new optimal values. The main advantage of SMO lies in the fact that solving for two Lagrange multipliers can be done analytically even though there are more optimization sub-problems that are being solved, each of these sub-problems can be solved fast and hence the overall QP problem is solved quickly.

3.4 Performance metrics

The Correlation Coefficient or coefficient of determination denoted R2 or r2 score which is used to moderation in the dependent variable means predicted from the independent variables. In this case, the r (CC) returns nearly 1.0 means a strong positive correlation. If the value of r returns nearly -1 means strong negative correlation and return 0 means no correlation between all the variables.

$$\boldsymbol{r} = \frac{\boldsymbol{n}(\sum \boldsymbol{xy}) - (\sum \boldsymbol{x})\,(\sum \boldsymbol{y})}{\sqrt{[\boldsymbol{n}\sum \boldsymbol{x^2} - (\sum \boldsymbol{x})^2] - [\boldsymbol{n}\sum \boldsymbol{y^2} - (\sum \boldsymbol{y})^2]}} \quad \ldots (3)$$

In the machine learning approach, Mean Absolute Error (MAE) means the average of absolute error in future prediction which means error range between prediction and observations. In data mining and ML research, the MAE denoted as loss function. The given accuracy formula is:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad \ldots (4)$$

where n is called the no. of elements in the iterations, Σ which is used to add them all up and $|y_i - x_i|$ called absolute error between actual and predicted.

The Root Mean Square Error (RMSE) is one of the familiar accuracy finding methods in data analysis, which is used to compute test the quality of prediction or forecasting. RMSE sometime name as root mean square deviation which is used to find the residuals between prediction and truth for all data points. The RMSE is calculated using the following formula.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \|y(i) - \hat{y}(i)\|^2}{n}} \quad \ldots (5)$$

where *n* is called the number of elements in the iterations, *y(i)* means i^{th} measurement, and *y ϒ(i)* called the prediction.

The Relative Absolute Error (RAE) is used to compute the accuracy for relatively comparison of each and every performance of a predictive model. The main reasons for calculating the RAE between actual and forecasted value. RAE is very useful to write the interpretation of the prediction which means if the RAE <1 means the model behavior is better. If RAE=0, the model behavior or accuracy is perfect.

$$RMSE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{\sum_{i=1}^{n} |y_i - \bar{y}|} \quad \ldots (6)$$

where *n* is called the number of elements in the observations, *y(i)* called the realized value and *y ϒ(i)* called the prediction and $\bar{y}$ means the mean values of corresponding variables.

Root Relative Squared Error (RRSE) is one of the accuracy metrics for predictive models called regression. It's an accuracy parameter which is used to compute the first result and behavior of model is performing. It is also an inheritance from RSE. The RRSE parameter for finding the process of square root for sum of squared errors for the corresponding predictive model with sum of squared errors.

$$RRSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad \ldots (7)$$

where *n* is called the number of elements in the observations, *y(i)* called the realized value and *y ϒ(i)* called the prediction and $\bar{y}$ means the mean values of corresponding variables.

4. Result and Discussion

Numerical illustrations based on Gaussian process and performance metrices using equations 1 to 7. The proof of proposed research and its numerical illustration is indicated in table 1 to table 7. The visualization based on tables indicate in figure 1 to figure 8. From experimental results, it can be found out that gaussian process, linear regression and sequential minimal optimization techniques under study can be used for micronutrients prediction.

Table 1 Statistical summary of micronutrients

Attributes	Max	Min	Mean	SD
Zn	1.7000	0.6000	1.1270	0.4070
Fe	22.4000	3.5000	10.2670	6.6610
Cu	2.6000	0.7360	1.4370	0.4900
Mn	3.3000	1.7400	2.5640	0.5080

Table 2 Distinct and uniqueness about micronutrients

Attributes	Distinct	Unique	Unique in %	Instances
Zn	25	19	59	32
Fe	27	23	72	32
Cu	30	28	88	32
Mn	22	15	47	32

Table 3 Performance of micronutrients parameters with R2 score

Attributes	Gaussian Processes	Linear Regression	Sequential minimal optimization
Zn	0.6594	0.7837	0.8100
Fe	0.7209	0.7771	0.7893
Cu	0.4987	0.5323	0.5430
Mn	0.0310	0.1629	0.4054

Table 4 Performance of micronutrients parameters with MAE

Attributes	Gaussian Processes	Linear Regression	Sequential minimal optimization
Zn	0.2887	0.1798	0.1772
Fe	4.7936	3.3354	3.3064
Cu	0.3496	0.3481	0.3129
Mn	0.4137	0.437	0.3374

Table 5 Performance of micronutrients parameters with RMSE

Attributes	Gaussian Processes	Linear Regression	Sequential minimal optimization
Zn	0.3375	0.251	0.2372
Fe	5.1419	4.2721	4.0745
Cu	0.439	0.4252	0.432
Mn	0.5161	0.5526	0.5264

Table 6 Performance metrices of micronutrients with RAE

Attributes	Gaussian Processes	Linear Regression	Sequential minimal optimization
Zn	76.5322	47.6618	46.9861
Fe	79.4931	55.3126	54.831
Cu	82.8982	82.5305	74.1946
Mn	91.6172	96.7787	74.7181

Table 7 Performance metrices of micronutrients with RRSE

Attributes	Gaussian Processes	Linear Regression	Sequential minimal optimization
Zn	81.4402	60.5657	57.2284
Fe	77.3131	64.2341	61.2637

Cu	90.0231	87.1973	88.6029
Mn	99.5067	106.5383	101.4926

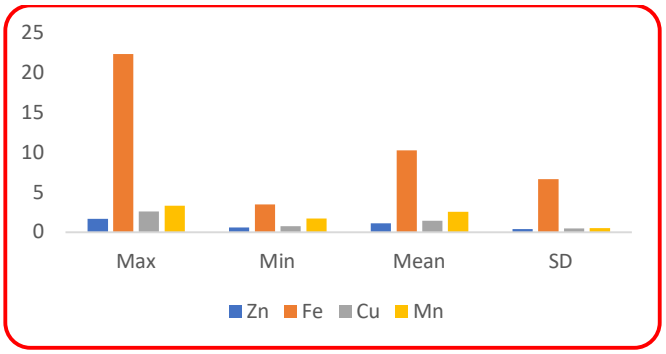

Fig. 2. Statistical summary of micronutrients

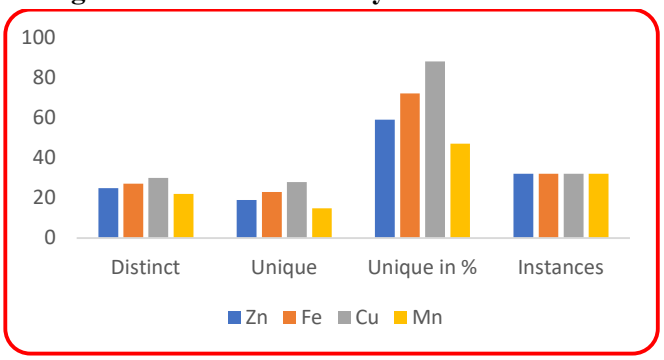

Fig. 3. Distinct and uniqueness about micronutrients

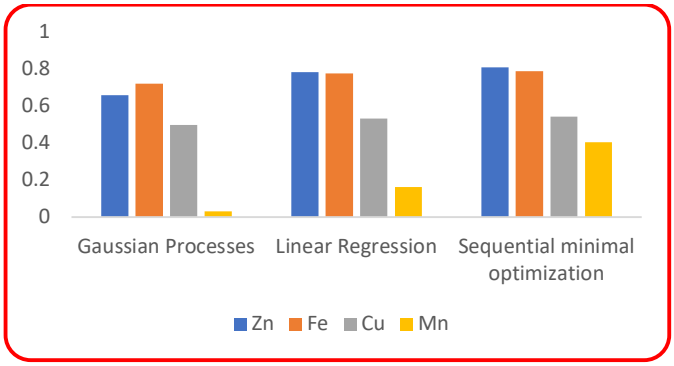

Fig. 4. Performance of micronutrients parameters with R2 score

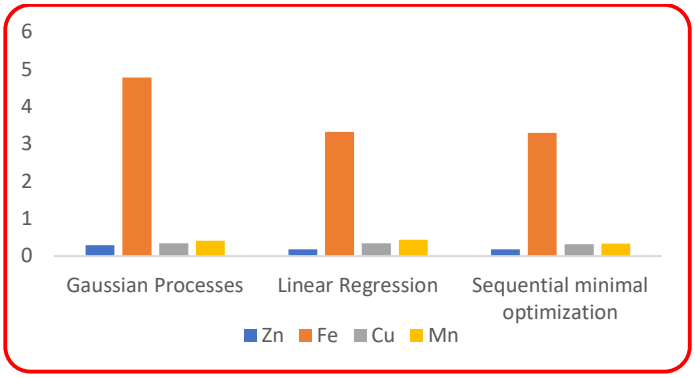

Fig. 5. Performance of micronutrients parameters with MAE

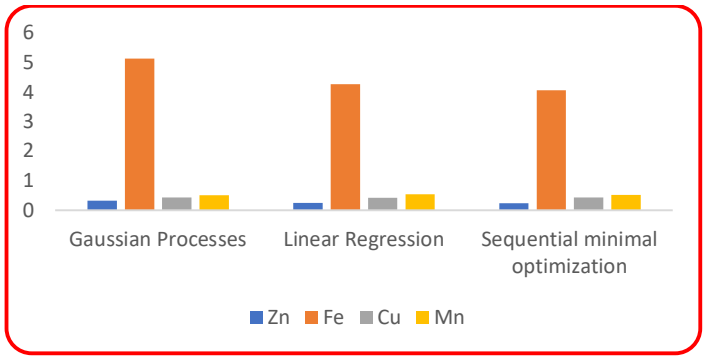

Fig. 6. Performance of micronutrients parameters with RMSE

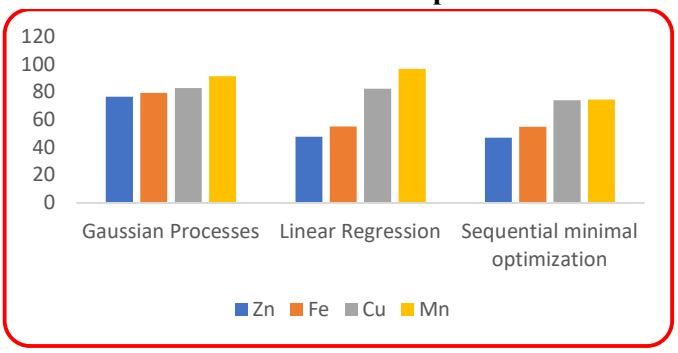

Fig. 7. Performance metrices of micronutrients with RAE

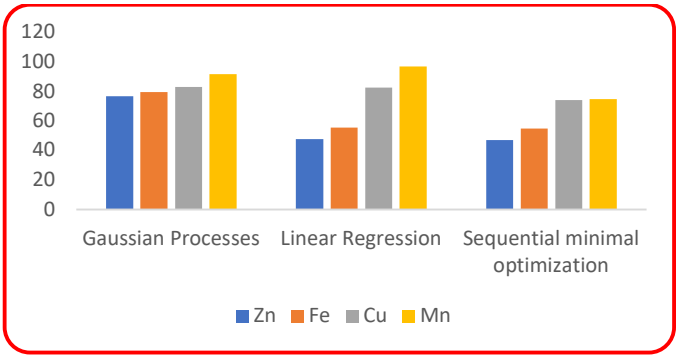

Fig. 8. Performance metrices of micronutrients with RRSE

Table 1 and figure 2 explain various statistical summaries of micronutrients level. This table indicates the mean and standard deviation of four parameters. The distinct and uniqueness of micronutrients are indicated in table 3 and figure 4.

The result and discussion are based on the performance of micronutrient parameters of Zn, Fe, Cu, and Mn using three different classifiers: the gaussian process, linear regression, and sequential minimal optimization. In this case, the SMO algorithm returns a Zn=81% strong positive correlation using all the nutrient levels. Furthermore, the linear regression produces a second level of positive correlation, and the gaussian process returns minimum correlation. the related numerical illustrations are shown in table 3 and figure 4.

Table 4 and figure 5 indicate the model performance of micronutrient parameters using MAE. In this test statistic, SMO returns a minimum error of 0.1772 (Zn) compared to other classification rules, and the root mean squared error also returns a minimum error. The related results are shown in table 5 and figure 6.

The performance metrics of micronutrients with Relative Absolute Error (RAE) and Root Relative Squared Error (RRSE) return SMO classifiers return minimum error compared to other classifiers. The related result is shown in table 5 and table 6. The numerical illustrations are based on graphs shown in figure 7 and figure 8.

Based on different machine learning classifiers for micronutrient analysis and prediction, the SMO algorithms return a strong positive correlation with minimum error compared to the other two classifiers, namely linear regression and the gaussian process.

5. Conclusion and Future Scope

From the experimental study, it can be concluded that Sequential Minimal Optimization (SMO) ML techniques can be effectively used for micronutrient data analysis and prediction. The effective ML techniques will help farmers predict yield in advance based on soil micronutrient parameters. In the future, Crop yield prediction with micro and macro nutrients datasets can be implemented using different machine learning classifiers, and decision tree approaches. Based on the results of nutrient prediction and recommendations can also be implemented to help farmers make decisions accordingly in case of low crop yield prediction.

References

1. Arokiyaraj, A, Juliat Mary, P., Vincentraj, A., & D.Sathya (2016). Macro and Micronutrient Status in Rice Growing Coastal Land Area of Tharangambadi Taluk of Nagapattinam District in Tamil Nadu, India, International Journal of Scientific & Engineering Research, 7(8), 69-75.
2. Awasthi, N. & Bansal, A. (2017). Application of Data Mining Classification Techniques on Soil Data using R. International Journal of Advances in Electronics and Computer Science, 4, 33-37.
3. Bhuyar, V. (2014). Comparative Analysis of Classification Techniques on Soil Data to Predict Fertility Rate for Aurangabad District. International Journal of Emerging Trends & Technology in Computer Science, 3(2), 200-203.
4. Brownlee, J. (2016). K Nearest Neighbors for Machine Learning. Retrieved March 23, 2020, from https://machinelearningmastery.com.
5. Chauhan, N. S. (2019). Introduction to Artificial Neural Networks (ANN). Retrieved March 26, 2020, from https://towardsdatascience.com.
6. Donges, N. (2019). A Complete Guide to the Random Forest Algorithm. Retrieved March 30, 2020, from https://builtin.com/data-science/random-forest-algorithm.html.
7. Gandhi, R. (2018). Naïve Bayes Classifier. Retrieved March 25, 2020, from https://towardsdatascience.com.
8. Hardesty. L. (2017). Explained: Neural Networks. Retrieved March 26, 2020, from https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414.
9. Harrison, O. (2018). Machine Learning Basics with the K-Nearest Neighbors Algorithm. Retrieved March 23, 2020, from https://towardsdatascience.com.
10. Jayalakshmi, R. & Devi, M. S. (2019). Relevance of Machine Learning Algorithms on Soil Fertility Prediction using R. International Journal of Computational Intelligence and Informatics, 8(4), 193-199.
11. Martin, K. Logistic Regression Models for Multinomial and Ordinal Variables. Retrieved March 27, 2020, from https://www.theanalysisfactor.com/logistic-regression-models-for -multinomial-and-ordinal-variables.html

12. Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press.
13. Nagesh, S. (2019). Real world implementation of Logistic Regression. Retrieved March 27, 2020, from http://towardsdatascience.com.
14. Priya, P., Muthaiah, U. & Balamurugan, M. (2018). Predicting Yield of the Crop Using Machine Learning Algorithm. International Journal of Engineering Sciences and Research Technology, 1-7.
15. Rajeshwari, V. & Arunesh, K. (2016). Analyzing Soil Data using Data Mining Classification Techniques. Indian Journal of Science and Technology, 9, 1-4.
16. Rakesh (2018). Performance analysis of rapeseed-mustard under different agro-climatic conditions of Jammu division of J&K state.
17. Rasmussen, C. E., & Williams, C. K. (2006). Gaussian processes for machine learning, 1, 159. Cambridge, MA: MIT press.
18. Renuka & Terdal, S. (2019). Evaluation of Machine Learning Algorithms for Crop Yield Prediction. International Journal of Engineering and Advanced Technology, 8(6), 4082-4086.
19. Shubham, J. (2018). Naïve Bayes Theorem. Retrieved March 25, 2020, from https://becominghuman.ai/naïve-bayes-theorem-d8854a41ea08.html.
20. Singh, V., Sarwar, A. & Sharma, V. (2017). Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach. International Journal of Advanced Research in Computer Science, 8(5), 1254-1259.
21. Smriti (2015). A Review on Soil Property Detection using Machine Learning Approach. International Journal Online of Science, 4.
22. Sujatha, R. (2016). A Study on Crop yield Forecasting using Classification Techniques. IEEE.
23. Supriya, D. M. (2017). Analysis of Soil Behavior and Prediction of Crop Yield using Data Mining approach. International Journal of Innovative Research in Computer and Communication Engineering, 5, 9648-9652.
24. Verma, A., Jatain, A. & Bajaj, S. (2018). Crop Yield Prediction of Wheat using Fuzzy C Means Clustering and Neural Network. International Journal of Applied Engineering Research, 13(11), 9816-9821.
25. Wang, J. (2020). An intuitive tutorial to Gaussian processes regression. arXiv preprint arXiv:2009.10862.
26. Yiu. T. (2019). Understanding Random Forest. Retrieved March 30, 2020, from https://towardsdatascience.com.
27. https://www.analyticsvidhya.com/blog/2021/05/all-you-need-to-know-about-your-first-machine-learning-model-linear-regression/
28. Jithin, (2020). https://bitmask93.github.io/ml-blog/Sequential-Minimal-Optimization-for-Support-Vector-Machines/