Applying Machine Learning Techniques to Extract dosages of Fertilizers for Precision Agriculture

P Singh¹, Ch Garg², and A Namdeo¹

¹Department of Civil Engineering, Amity University Uttar Pradesh, Noida, India
²Department of Computer Science and Engineering, Amity University Uttar Pradesh, Noida, India

*Email: priyanka24978@gmail.com

Abstract. Increasing cost of fertilizers and subsidies are proving to be a burden on agriculturists as well as the government. Now it’s high time to involve technology so that precise amount of fertilizers could be added in order to obtain maximum yield. Machine learning is an important tool that can be used to predict precise nitrogen, phosphorous and potassium for fertigation. But due to different soil types and different needs of different crops and varieties, it has become difficult to predict the exact amount of fertilizer needed. In this research paper a solution has been drawn to obtain precision in agriculture. Data has been collected from soil reports, soil science institutes and agriculture institutes of India. It has been rearranged into tables and pre-processed for applying machine learning models. Once the models were applied their accuracies have been evaluated using various parameters. Also, the predictions given by our model were compared by already existing recommendations. This work has been done for irrigated wheat growing areas of India and it could be extended to other crops and other areas all over the world.

1. Introduction
According to FAO, world population will reach to 9.1 billion by 2050. To cater the food requirements of such a large population there will be a need to increase the agricultural production by 70 percent [1]. The current scenario of growing demand for food products may lead to intensive farming [2] and hence, environmental crisis [3, 4]. The consumption of nitrogen fertilizer alone has amounted to 100 million tonnes but only 33 percent of applied nitrogen is used by the plants, rest 67 percent is lost in the environment [5]. This nitrogen may be released as a pollutant in ground water or may be emitted as nitrous oxide into the atmosphere [6]. Also phosphorous from farms, enter into water systems. This is a major cause of algal blooms and ultimately dead zones [7, 8]. So there arises a dire need to save the environment from agricultural pollutants.

We can reduce the hazard caused to the nature due to agricultural practices by implementing precision agriculture. One of the methods that can be used to apply precision agriculture is predicting the exact amount of fertilisers to be added in a farmland. Earlier fertiliser was added into the farms as per the blanket recommendations suggested by the agriculturists for each crop. For example, the blanket recommendation for indigenous varieties of wheat was 80:40:40 for N:P:K. Now a days, a new approach is used. The entire nation is divided into zones and the fertilisers for each zone are recommended crop wise and category wise. By categories we mean the ranges for the amounts of nutrients present in the soil.
They are categorised as very low, low, medium, high and very high. Figure 1 and figure 2 show the dosage recommendation for the crop wheat in the farmlands of Haryana. These figures clearly exhibit that the amount of nitrogen present in the soil vary by almost 100 kg/ha but the dosage recommended remains the same. Moreover, there is no mention of the target yield. So in many cases these recommendations may not prove to be cost effective.

**Figure 1. Recommended Dosage for Crops [9]**

Our research work has focused on applying machine learning techniques to provide precise recommendations of fertilisers, for each and every farmland individually. We have also considered target yield as one of the input parameters. This would help the farmers to make financial decisions as well.
2. Method
Machine learning is a technique through which systems can learn from their experiences and improve without being programmed explicitly. It is a technology that emphasises on learning from data. The machine identifies the different patterns formed in the datasets. Then, based on these analysis, predictions are made.

The process of machine learning includes data collection, data preparation, training the machine and evaluation of the model.

2.1. Data Collection and Data Preparation
Wheat is a staple crop in the northern parts of India. So, for this research work we focused on irrigated wheat growing areas of India. We had collected data from soilhealth.dac.gov.in and iiss.nic.in. These are the portals created by the Ministry of Agriculture and Farmer Welfare, Government of India and Indian Institute of Soil Sciences, Bhopal respectively. Both these datasets are a benchmark used by the agriculturists. Once we had retracted the data from these websites, we organised and assimilated the information into a single dataset. This was done so that we can apply the machine learning techniques efficiently for prediction. The features that were included in our dataset were zone, pH, electrical conductivity (EC), organic carbon content (OC), soil nitrogen, soil phosphorus, soil potassium, recommended nitrogen, recommended phosphorus and recommended potassium.

---

**Figure 2. Recommended Dosage for Crops [9]**

| Parameter                  | Value | Unit |
|----------------------------|-------|------|
| Organic Carbon (OC)        | 0.05  | %    |
| Available Nitrogen (N)     | 12.75 | kg/ha|
| Available Phosphorus (P)   | 11.76 | kg/ha|
| Available Potassium (K)    | 134   | kg/ha|

**Fertilizer Recommendations**

| Fertilizer Combination1          | Value | Fertilizer Combination2          | Value |
|----------------------------------|-------|----------------------------------|-------|
| Neem Coated Urea                 | 454.35| Diammonium Phosphate (16:44:0)   | 120.45|
| Single Superphosphate (16% P2O5 | 331.25| Neem Coated Urea                 | 412.45|
| Granulated)                      |       | Potassium Chloride (Muriate of Potash) | 66.67|
**Table 1. Focus Areas**

| Zones              | Districts |
|--------------------|-----------|
| Bihar              | Rohtas, Bhojpur, Buxar, Bhabhua, Arwal, Patna, Nalanda, Nawadah, Jehanabad, Aurangabad, Gaya |
| Haryana            | All districts |
| Himachal Pradesh   | Kangra, Sirmour, Madi, Kullu, Shimla |
| Jharkhand          | All districts |
| Madhya Pradesh     | Bhopal, Dhar, Jabalpur, Indore, Khandwa, Khargone, Mandsaur, Narsinghpur, Powarkheda, Rewa, Satna, Sagar, Sehore, Ujjain |
| Maharashtra        | Ahmednagar, Pune, Jalgaon, Nasik, Aurangabad, Parbhani, Julna, Akola, Buldhana, Wardha, Yawatmal, Satara, Sangli, Kolhapur, Dhule, Nandurbar |
| Uttar Pradesh      | Gautam Budha Nagar, Ghaziabad, Bagpat, Meerut, Mujjafarnagar, Saharanpur, Bulandshahr, Aligarh, Mahamayanagar, Etah, Agra, Etawah, Mainpuri, Shikohabad, Mathura, Jhansi, Firozabad, Jalaun |
| West Bengal        | Nadia, Burdwan, Murshidabad, 24 Paraganas North districts |

**Table 2. Details of Input and Output Parameters**

| Input Parameter | Input |
|-----------------|-------|
| X1              | Zone  |
| X2              | pH    |
| X3              | Electrical Conductivity |
| X4              | Organic Carbon |
| X5              | Soil Nitrogen |
| X6              | Soil Phosphorus |
| X7              | Soil Potassium |

| Output Parameter | Output |
|------------------|--------|
| Y1               | Recommended Nitrogen |
| Y2               | Recommended Phosphorus |
| Y3               | Recommended Potassium |

2.2. *Training the Machine and Model Evaluation*

Once the data was assimilated, we used Minitab as a tool to apply the machine learning techniques. We trained the machine using different regression models. Thereafter, we evaluated all the models on the basis of their accuracies. CART regression was found to be the best fit model. We established three different models for making three different recommendations of macronutrients nitrogen, phosphorus and potassium. While training the machine we used Zones as a categorical parameter and the remaining inputs as continuous parameters for predicting the fertiliser dosages.

**Method**

- **Node splitting**: Least squared error
- **Optimal tree**: Within 1 standard error of maximum R-squared
- **Model validation**: 10-fold cross-validation
- **Rows used**: 1682

**Figure 3.** Approach to apply CART regression
The method that we applied to make predictions for all the three models using CART regression is mentioned in figure 3. We used least squared error for node splitting and we applied 10 fold cross validation to avoid overfitting. For evaluating the model we used various parameters like R squared value and root mean squared error. or ideal condition, R - squared value should be 1 and RMSE value should be 0.

3. Results

3.1. CART Regression Model for Recommendation of Nitrogen

The model that was applied for recommending nitrogen used 138 terminal nodes. It was more than 90 percent accurate as depicted by the R - squared and the RMSE values in figure 4. The figure also shows that all the input parameters are important for making predictions for the fertiliser dosage of nitrogen. Figure 6 and figure 7 show the goodness of fit of our regression model.

| Model Summary |
|----------------|
| Total predictors | 8 |
| Important predictors | 8 |
| Number of terminal nodes | 138 |
| Minimum terminal node size | 3 |
| Statistics | Training | Test |
| R-squared | 0.9951 | 0.9899 |
| Root mean squared error (RMSE) | 7.0545 | 10.0775 |
| Mean squared error (MSE) | 49.7655 | 101.5550 |
| Mean absolute deviation (MAD) | 4.3288 | 5.9251 |
| Mean absolute percent error (MAPE) | 0.0884 | 0.1067 |

Figure 4. Model Summary for Nitrogen Recommendation
3.2. **CART Regression Model for Recommendation of Phosphorus**

A 134 node model was obtained after applying the CART regression for recommending phosphorus. As shown in figure 8, the model applied was about 95 percent accurate. Also the scatterplot of response fit versus actual fit and the residual plot of terminal nodes showed the extent of similarity in the recommended value of phosphorus and the predicted value given by the model. This shows that are model can be used to make precise predictions for phosphorus.
**Model Summary**

Total predictors: 8  
Important predictors: 8  
Number of terminal nodes: 134  
Minimum terminal node size: 3

| Statistics                                | Training  | Test  |
|-------------------------------------------|-----------|-------|
| R-squared                                 | 0.9950    | 0.9872|
| Root mean squared error (RMSE)            | 3.1410    | 5.0354|
| Mean squared error (MSE)                  | 9.8660    | 25.3550|
| Mean absolute deviation (MAD)             | 1.9587    | 2.8194|
| Mean absolute percent error (MAPE)        | 0.0416    | 0.0640|

**Figure 8.** Model Summary for Phosphorus Recommendation

**Figure 9.** Relative Variable Importance for Phosphorus Recommendation

**Figure 10.** Scatter Plot of Response Fit vs Actual Fit for Phosphorus Recommendation
3.3. CART Regression Model for Recommendation of Potassium

As depicted in the model summary, a decision tree of 106 nodes was used to precisely predict the amount of potassium to be added in the fields. All the 8 predictors were used to devise a 95 percent accurate model. The relative variable importance, the scatterplots and the residual plots showed that our model was a good fit.

**Figure 12. Model Summary for Potassium Recommendation**

|                         | Training | Test  |
|-------------------------|----------|-------|
| Total predictors        | 8        |       |
| Important predictors    | 8        |       |
| Number of terminal nodes| 106      |       |
| Minimum terminal node size | 3       |       |
| R-squared               | 0.9968   | 0.9893|
| Root mean squared error (RMSE) | 3.0436 | 5.5498|
| Mean squared error (MSE) | 9.2633  | 30.8007|
| Mean absolute deviation (MAD) | 1.7707 | 2.7252|
| Mean absolute percent error (MAPE) | 0.0706 | 0.1331|
Figure 13. Relative Variable Importance for Potassium Recommendation

Figure 14. Scatter Plot of Response Fit vs Actual Fit for Potassium Recommendation
When we evaluated these models for accuracies, we also checked if the model is overfitting. We found that there was a minimal difference in the R squared values for training and testing data. This testified that our models were neither overfitting the data nor was it underfitting.

We also observed that zone is the most important parameter in all the three cases. This pattern is justified since each zone follows similar agricultural practices, has similar terrains, similar parent rock and almost similar climatic conditions. Therefore, there are different recommendations for different zones.

### 4. Discussions

As mentioned earlier there were similar recommendations made by the agricultural institutes of India for fertiliser dosages even for adversely varying quantities of soil nitrogen. For example, 209 kg/ha nitrogen was recommended for both the soil nitrogen values 12.75 kg/ha and 109.65 kg/ha. Moreover, target yield was not taken into consideration for making these recommendations. Same quantities of nitrogen were suggested for different target yields.

The model proposed by us helps to predict the precise values of fertilisers to be added in the farmland for different values of soil nutrients and different target yields. The predictions made by our model for the zone Haryana are mentioned in figures 16, 17, 18 and 19. Figure 16 shows that if the farmland contains 12.75 kg/ha of soil nitrogen and the required yield is 50 kg/ha, the recommended dosage for nitrogen is 266.694 kg/ha. But when the target yield changes to 45 kg/ha, the amount of nitrogen to be added in the fields is predicted to be 237.848 kg/ha as shown in figure 17. The recommendations also show a significant change when the soil nitrogen values change. In figure 18 and 19 the amount of nitrogen present in the field is 109.65 kg/ha. The suggested dosage of nitrogen for the target yield 50 kg/ha and 45 kg/ha is 140.908 kg/ha and 114.046 kg/ha respectively. This proves that our model shows diverse recommendations for varied conditions.
Figure 16. Recommended Nitrogen for Soil N = 12.75 kg/ha, Target Yield = 50 kg/ha

Prediction for RN

Settings

\[ \text{pH} = 7.64, \text{EC} = 0.43, \text{OC} = 0.05, N = 12.75, P = 11.76, K = 134, \text{Target Yield} = 50, \text{Zone} = \text{haryana} \]

Prediction

| Terminal | Obs | Node ID | Fit  |
|----------|-----|---------|------|
| 1        | 66  | 266.694 |

Figure 17. Recommended Nitrogen for Soil N = 12.75 kg/ha, Target Yield = 45 kg/ha

Prediction for RN

Settings

\[ \text{pH} = 7.64, \text{EC} = 0.43, \text{OC} = 0.05, N = 12.75, P = 11.76, K = 134, \text{Target Yield} = 45, \text{Zone} = \text{haryana} \]

Prediction

| Terminal | Obs | Node ID | Fit  |
|----------|-----|---------|------|
| 1        | 46  | 237.848 |

Figure 18. Recommended Nitrogen for Soil N = 109.65 kg/ha, Target Yield = 50 kg/ha

Prediction for RN

Settings

\[ \text{pH} = 7.31, \text{EC} = 0.23, \text{OC} = 0.43, N = 109.65, P = 8.58, K = 130, \text{Target Yield} = 50, \text{Zone} = \text{haryana} \]

Prediction

| Terminal | Obs | Node ID | Fit  |
|----------|-----|---------|------|
| 1        | 81  | 140.908 |

5. Conclusions

Three different models have been prepared using CART regression to predict the amount of nitrogen, phosphorus and potassium to be added in the field. The models proposed in this research work are more than 90 percent accurate and consider all the eight parameters that are zone, pH, EC, OC, soil N, soil P, soil K and target yield to be important for making recommendations. The research work clearly shows that the prediction for fertiliser dosage are made according to each and every farmland, individually. Also, target yield taken as a parameter for suggesting fertiliser dosage will help the farmers to take their financial decisions precisely. This fulfils our objective of being precise while suggesting the amount of nutrients to be added in a field. This research work has been done for irrigated wheat growing areas. The same model can be used for other crops like rice, maize, pulses, cash crops, etc. and for other areas all around the globe.

References

[1] FAO 2009, How to Feed the World in 2050’, Food and Agriculture Organization of the United Nations, Rome.
[2] Matson PA, Parton WJ, Power AG, Swift MJ 1997 Science 277 504-509.
[3] Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C 2010 Science 327 812-818.
[4] Tilman D, Balzer C, Hill J, Befort BL 2011 Proceedings of the National Academy of Sciences of the United States of America 108 20260-20264.
[5] López-Bellido RJ, López-Bellido L 2001 Field Crops Research. 71 31-46.
[6] Praba R, Singh M, Singh AB 2005 Indian Journal of Agricultural Sciences 75 823-825.
[7] Mallarino AP, Haq MU 2015 Proceedings of the North Central Extension-Industry Soil Fertility Conference, Des Moines, Int. Plant Nutr. Inst., Peachtree Corners, GA, pp 94-100.
[8] Singh B, Singh Y, Sekhon GS 1995 Int. J.Env.Sci. 20(2) 167-184.
[9] www.https://soilhealth.dac.gov.in/

Figure 19. Recommended Nitrogen for Soil N = 109.65 kg/ha, Target Yield = 45 kg/ha