International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 03 Issue: 08 | August -2019 ISSN: 2590-1892

# FERTILIZER PREDICTION USING PYTHON AND MACHINE LEARNING MODELS THROUGH BIGDATA ANALYTICS

**KAVERI MIRJI[1], HARSHITHA H J[2], VISHWANATH P L[3], SANJAY K T[4], ABHISHEK PATIL[5], MAHENDRA KUMAR B[6]**

*[1, 2,3,4,5] M.C.A. Post Graduate Students D.S.C.E*
*[6]CO-AUTHOR M.C.A Assistant Professor D.S.C.E*

---

**Abstract:**

Agriculture plays a crucial role in the life of an economy. It is the backbone [16] for developing countries like India as more than 75% of population depends on agriculture. To increase crop production many factors are responsible like soil, weather, rain, fertilizers and pesticides.[17]To maintain nutrient levels in the soil in case of deficiency, fertilizers are added to soil.[14]

The common problem existing among the Indian farmers is that they choose approximate amount of fertilizers and add them manually. Excess or insufficient addition of fertilizer can harm the plant life and reduce the yield.[14] This paper provides review of various data mining techniques used on agriculture soil dataset for fertilizer prediction. Mainly we focused on various parameters like Ca, Mg, K, S, N, Lime, C, P and Moisture value etc.

**Keywords**: Agriculture, *Fertilizer, Big Data, Classification, Machine Learning, logit, Regression.*

## 1. INTRODUCTION

A **fertilizer** is any material of natural or synthetic origin that is applied to soils or to plant tissues to supply one or more plant nutrients essential to the growth of plants. Many sources of fertilizer exist, both natural and industrially produced.[13]

Fertilizers enhance the growth of plants. This goal is met in two ways, the traditional one being additives that provide nutrients. The second mode by which some fertilizers act is to enhance the effectiveness of the soil by modifying its water retention and aeration. This article, like many on fertilizers, emphasizes the nutritional aspect. Fertilizers typically provide, in varying proportions:[11]

- three main macronutrients:
  - Nitrogen (N): leaf growth
  - Phosphorus (P): Development of roots, flowers, seeds, fruit;
  - Potassium (K): Strong stem growth, movement of water in plants, promotion of flowering and fruiting;
- Three secondary macronutrients: calcium (Ca), magnesium (Mg), and sulfur (S);
- micronutrients: copper (Cu), iron (Fe), ma nganese (Mn), molybdenum (Mo), zinc (Z n), boron (B). Of occasional significance are silicon (Si), cobalt (Co), and vanadium (V).[11]

The nutrients required for healthy plant life are classified according to the elements, but the elements are not used as fertilizers. Instead compounds containing these elements are the basis of fertilizers. The macro-nutrients are consumed in larger quantities and are present in plant tissue in quantities from 0.15% to 6.0% on a dry matter (DM) (0% moisture) basis. [11]

## 2. BODY OF THE PAPER

Predictive analytics includes various techniques from machine learning, data mining techniques that are derived from different historical information and currents factors to make smarter decision about the future events. Predictive analytics includes various machine learning techniques and

statistical techniques for classification and pattern matching. This includes analyzing what event had occurred in the past, when it is occurred, and monitoring current terminologies and facts about data, then making smarter decision using machine learning methods.

In Prediction, previous year data are referred as training set and the data's are classified based on training set .[15]The first and foremost process in prediction is data collection then data analysis, monitoring and finally result can be plotted as graph using matplotlib function.Broadly, there are 3 types of Machine Learning Algorithms:

### A. Supervised Learning

This algorithm consist of a target / outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). Using these set of variables, we generate a function that map inputs to desired outputs. The training process continues until the model achieves a desired level of accuracy on the training data. Examples of Supervised Learning: Regression, Decision Tree, Random Forest, KNN, Logistic Regression etc.[10]

### B. Unsupervised Learning

In this algorithm, we do not have any target or outcome variable to predict / estimate. It is used for clustering population in different groups, which is widely used for segmenting customers in different groups for specific intervention. Examples of Unsupervised Learning: Apriori algorithm, K-means.[10]

### C. Reinforcement Learning:

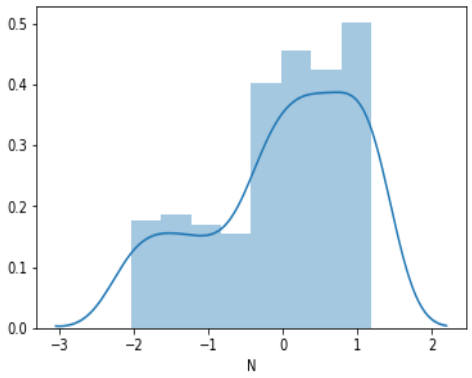
Using this algorithm, the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it trains itself continually using trial and error. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions. Example of Reinforcement Learning: Markov Decision Process.[10]

### 2.1 Logistic Regression Model

It is used to estimate discrete values (Binary values like 0/1, yes/no, true/false) based on given set of independent variable(s). In simple words, it predicts the probability of occurrence of an event by fitting data to a logit function. Hence, it is also known as **logit regression**. Since, it predicts the probability, its output values lies between 0 and 1 (as expected).[10]

The below distplot figure displays the estimation of predicted values using Logistic Regression

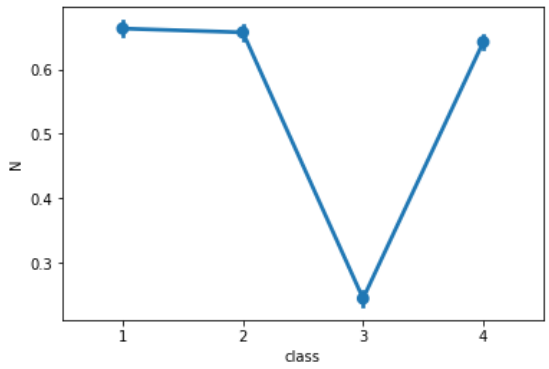
Fig 1 – Output of Logistic Regression Model




### 2.2 Decision tree

It is a type of supervised learning algorithm that is mostly used for classification problems. Surprisingly, it works for both categorical and continuous dependent variables. In this algorithm, we split the population into two or more homogeneous sets. This is done based on most significant attributes/ independent variables to make as distinct groups as possible.[12]

Fig 2 – Pointplot graph




### 2.3 Random Forest

Random Forest is a trademark term for an ensemble of decision trees. In Random Forest, we've collection of decision trees (so known as "Forest"). To classify a new object based on attributes, each tree gives a classification and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest).[12]

Each tree is planted & grown as follows:

1. If the number of cases in the training set is N, then sample of N cases is taken at random but *with replacement*. This sample will be the training set for growing the tree.
2. If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.
3. Each tree is grown to the largest extent possible. There is no pruning.[12]

## CONCLUSION

The attempts are made to predict the consumption of fertilizers to increase the crop yield through various machine learning algorithms such as classification and regression models. The best fertilizer based on yield rate are being researched for the problem. In our future work, we are including soil parameters, temperature, and rainfall for more accurate results.

## REFERENCES

[1]https://www.irjet.net/archives/V4/i5/IRJET-V4I5513.pdf

[2] https://www.britannica.com/topic/fertilizer

[3] https://adalidda.com/posts/2fZRnWsEc4vmrJuxs/fertilizer-prediction-through-big-data-analytics

[4] https://en.wikipedia.org/wiki/Fertilizer

[5] https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/

[6] https://pdfs.semanticscholar.org/df07/bf679bc62d68963f871d4feaeffaaf54f566.pdf

[7] https://www.irjet.net/archives/V5/i2/IRJET-V5I2479.pdf

[8]https://www.mdpi.com/20751702/6/3/38/pdf

[9] https://www.chemicalsafetyfacts.org/fertilizers/

[10] Referred website:https://madoveranalytics.com/

[11] Referred website:

https://en.wikipedia.org/

[12] Referred journal paper published by

http://www.unipune.ac.in/administration_files/publication.htm - SavithriBai Phule Pune University

[13] Referred technical paper by

https://www.researchgate.net/institution/Universiti_Teknologi_MARA2 - Universiti Teknologi MARA

[14] Referred web resource- http://acadpubl.eu/ap/

[15] Referred article-Classification of rubberized coir fibres using deep learning-based neural fuzzy decision tree approach by S.Remya and Sasikala Ra.

[16] Referred journal paper by http://thegrenze.com/

[17] Referred website https://www.ijettcs.org/

[18] Referred we resource http://ijcsse.org/