PREDICTION AND ANALYSIS OF SOIL MACRONUTRIENTS USING MACHINE LEARNING TECHNIQUES

Sharmila A[1], Gokul Kannan S P[2], Abhinanthan S S[3], Nithish Kumar S[4]

[1] *Associate Professor, Electronics and Communication Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India*

[2] *Student, Electronics And Communication Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India*

[3] *Student, Electronics And Communication Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India*

[4] *Student, Electronics And Communication Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India [1]*

ABSTRACT

Everyone has to be healthy and have access to enough crops and food as the world's population rises. The nation's economy benefits from crop output. Accurately estimating the percentage of macronutrients in the soil will assist in selecting the best crop to cultivate. Recent years have seen numerous advancements in everything from harvesting to product selection. The yield is increased when the correct crop is chosen for growing. Understanding the requirements for macronutrients is crucial for achieving optimal yield. The amount of NPK levels needed for various crops varies. The variables that must be taken into account are pH, temperature, humidity, and rainfall. Different machine learning models are often trained, tested, and validated. Prediction techniques include Decision Trees, AdaBoost Classifiers, XGB Classifiers, Random Forests, Logistic Regressions, SVM (Support Vector Machine) Classifiers, and KNN Algorithms. We can select the most effective model by contrasting the accuracy of several models. The creation of a user-friendly website is the suggested process. Inputs from the user include NPK levels, pH, temperature, humidity, and rainfall. Following the machine learning model's study, the website will take the macronutrients values as the input and suggest us the appropriate crop that may be cultivated in a given area as an output. We saw improved accuracy in the KNN Algorithm, SVM Classifier, and logistic regression during training, testing, and validation. The KNN algorithm's accuracy was determined to be 0.9886. We have used Google colab for training and testing the machine with datasets. We have created a user-friendly website by merging these two procedures.

Keyword: - *Soil Macronutrients, NPK values, KNN, SVM classifier.*

---

1. INTRODUCTION

A farmer needs to understand the properties of the soil in order to be successful. The foundation of health is the soil. In India, agriculture must become more sustainable in order to fulfill the country's expanding needs. In order to feed the world's 1.3 billion inhabitants, agriculture occupies more than 60% of the country's land. Any nation's economy is reliant on its agriculture sector. Factors including crop demand, weather, soil analysis, and water availability are crucial in agriculture. The biggest issue facing agriculture is ignorance of climate fluctuations. Sensors or laboratory soil testing can be used to identify the minerals in the soil. Numerous sensors can be used to gather data in real time. From planting seeds to harvesting the produce, several innovations have been accomplished in the modern era. Cultivators can benefit from IoT, cloud computing, and machine learning at every stage of the cultivation process. The steps involved in agriculture can be broadly categorized as selecting appropriate seeds, keeping an eye on crop growth, guarding against illnesses, determining when the crop is ready to be harvested, and so on. A support structure ought to be created in a way that enables farmers to make quick decisions and lessens their reliance on the regional agricultural offices. The amount of macro- and micronutrients in the soil affects how much crop production is produced. A number of soil characteristics must be examined before deciding which crop to cultivate. The ratio of nitrogen (N), phosphorus (P), and potassium (K) in the soil determines how much a plant can grow.
Any crop's ability to grow is influenced by a variety of elements, including the quality of the soil, the climate, the availability of water, the temperature of the soil, air humidity, sunlight, wind, and pollution levels. The features of the soil and the availability of minerals change periodically. Farmers may apply the incorrect amount of fertilizer without knowing the precise NPK values, which lowers crop output. The relationship between soil pH and NPK is vital. Pattern recognition can be achieved by machine learning (ML).Machine learning algorithms utilize mathematical or statistical models for analysis and prediction. Within the field of artificial intelligence, machine learning employs past data as a means of forecasting future values. Based on the correlation between dependent and independent variables, machine learning makes predictions. Because soil pH, moisture content, temperature, and cation exchange capacity are independent factors, nitrogen, phosphorus, and potassium are dependent variables. More independent variables lead to better results from machine learning. Machine learning can be used to predict crops that are in demand by utilizing market data and past weather trends. Reinforcement learning, unsupervised learning, and supervised learning are the three main categories of machine learning. Eighty percent of the dataset is utilized for testing, and the remaining twenty percent is used for training. This divides the dataset in two. The machine learning system is trained with sample-labeled data in supervised learning.

1.2 ADVANTAGES:

The use of machine learning algorithms to predict and analyze soil macronutrients is revolutionizing agricultural processes and has many benefits. First off, machine learning models outperform conventional techniques in terms of precision and accuracy, enabling more trustworthy predictions based on complex data interactions. Because of the increased precision, decision-making is improved and becomes more efficient, saving time and resources. Furthermore, even with the initial setup costs, machine learning proves to be cost-effective in the long run because it saves labor costs and optimized resource allocation. A major benefit is environmental sustainability, since precise forecasts minimize fertilizer usage, lower pollution levels in the environment, and enhance soil health over the long run. Machine learning helps make the agricultural industry more sustainable and profitable by advancing sustainable farming methods and precision agriculture. Predictive models give farmers, agronomists, and legislators useful information to help them make decisions about crop selection, fertilizer use, and soil management techniques. Furthermore, machine learning algorithms are more applicable in a variety of agricultural contexts when they are tailored to certain soil types, climates, and crop kinds.

1.3 APPLICATIONS:

Machine learning has numerous and significant applications in a wide range of industries for the analysis and prediction of soil macronutrients. These methods provide farmers with accurate information about the health of their soil, which helps them choose crops and apply fertilizer more wisely for increased yield and sustainability. Continuous monitoring of soil health is beneficial to environmental scientists as it facilitates the early detection of nutrient deficits and provides information for repair operations. Furthermore, planners may create focused plans for minimizing environmental effects and maximizing land use by using assessments powered by machine intelligence. The spatial mapping of soil nutrient levels is further improved by integration with remote sensing and GIS, enabling scaled-up, well-informed decision-making.

Machine learning models help precision farmers monitor and control soil nutrient levels in real time. This allows for site-specific modifications to fertilization and irrigation schedules, which reduces waste and maximizes resource utilization. This accuracy lessens the impact on the environment while improving crop health and output yield of the crops .In order to create thorough predictive models, machine learning techniques allow the integration of several data sources, such as weather patterns, crop growth phases, and previous soil data. Future trends in soil nutrients can be predicted by these models, which helps with long-term planning and decision-making in land management and agriculture.

Applications of machine learning to the prediction and analysis of soil macronutrients are found in a wide range of fields, including environmental monitoring, precision agriculture, sustainable land management, and global food security programs. With its ongoing improvement and integration with other technologies, machine learning has the potential to significantly influence how agriculture and land stewardship are conducted in the future.

2. LITERATURE SURVEY

Angu raj et al., developed a system using IoT (Internet of Things) and machine learning. Soil parameters are collected by sensors and fed into the training model. The machine learning models used are Naive Bayes and Random Forest. Naive Bayes classifies objects based on their probability and thus conforms to the required classification. Naive Bayes can be used for binary distributions as well as many other types of distributions. Bayes theorem is based on the principle of Bayes rule. Bayes' law depends on probability. The results of Bayes' rule can also be explained as posterior probability. The posterior probability can be defined as the ratio of the probability multiplied by the prior probability to the relevant probability.

Murali krishna senapathy et.al., proposed an IoT enabled soil nutrient classification and crop recommendation model. In their model, they recorded humidity, temperature, NPK and pH using IoT sensors. They used the Fruit Fly Optimization (FFO) algorithm to optimize model selection for multivariate support vector machine (MVSM) models. Used for soil and water samples LDR color sensor, soil moisture sensor, pH sensor, thermometer, hydrometer, and GPS. Models such as linear SVM, kernel SVM and decision trees are used. The combination of MVSM and FFO improves classification accuracy.

Ersin elbasi et.al., in their paper explored and discussed various classification algorithms. The objective of the study is to incorporate IoT sensors and ML algorithms for waste reduction and reap production optimization. The performance of machine learning algorithms such as SVM, Naïve Bayes, Decision tree, and Random Forest were analyzed. The study identified and presented an experimental result of the opportunities and challenges in integration of IoT Sensors and ML models. Appropriate feature selection becomes an important factor to achieve better accuracy. Temperature, humidity, pH, precipitation, etc. factors serve as criteria. Using Naive Bayes, the accuracy is 97.05%. The accuracy of random forest is 97.32%.

Amna Ikram introduced a model that uses IoT devices and machine learning (ML) algorithms to maximize crop yields, helping farmers increase profitability. The main factors in crop selection are nutrients such as nitrogen, phosphorus, potassium, and CO_2, as well as soil conditions such as soil temperature, moisture, and rainfall. Data for the above features is collected from real-time audit data and sent to Firebase Cloud for analysis. Five machine learning algorithms namely Decision Tree, SVM, KNN, Random Forest and Gaussian Naive Bayes are used to train the model to improve performance and accuracy.

3. OBJECTIVE AND METHODOLOGY

3.1 OBJECTIVES

To evaluate the macronutrients content of the soil and forecast crops that are suitable for growing in a certain area based on (such nitrogen, phosphorus, and potassium), temperature, humidity, pH, and rainfall in the field of agriculture. It is common practice to train, test, and evaluate different machine learning algorithms in order to choose an appropriate machine learning model for the system. By contrasting the performance of different algorithms, the optimal algorithm may be selected.

Google colab is used to train the machine with different datatypes. To create a website that accepts values in input fields and outputs data that corresponds to the input.

3.2 METHODOLOGY

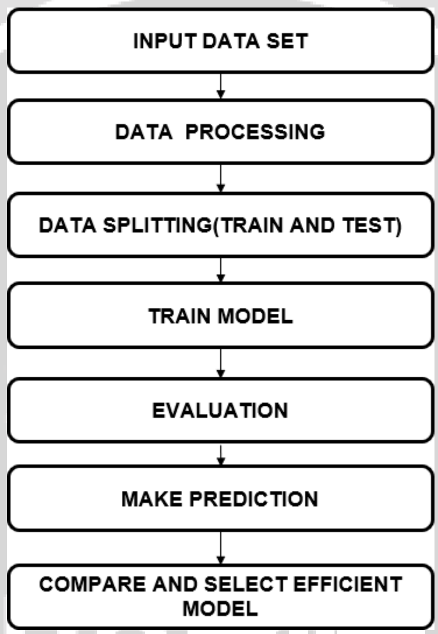

Figure 3. Proposed methodology for Soil Macronutrients Prediction.

The process of teaching a machine learning algorithm to predict target variables from a set of inputs is known as the training process in machine learning. In this process, input is fed into the algorithm, the values are compared, and the algorithm's parameters are adjusted until it can predict the goal value for the input features.

The dataset is being trained, tested, and validated using a variety of machine learning methods, including

1. Decision tree
2. Classifier AdaBoost
3. The XGB classification

4. A woodland at random
5. Regression with logs
6. Classifier SVM
7. KNN algorithm; 8. Recognition of patterns in neural networks

All of these algorithms fall into two categories: classification and regression.

3.2.1 Classification:

Sorting input data into different classes or labels according to its features is the aim of this supervised learning activity. The input properties of the labeled training data, which correspond to predetermined output classes, are used to teach the algorithm. Identifying new, unseen data points and correctly predicting their class label is the aim of a classification algorithm.

- **Supervised learning**, One kind of machine learning is called supervised learning, where algorithms are taught using recorded data—that is, input and output data together. By searching for patterns in the domain, the algorithm finds patterns during training and learns to match input data with the correct output. After being trained, a model can decide or make predictions based on fresh, unobserved data.
- **Unsupervised learning,** That being said, unsupervised learning operates on untagged data. In unsupervised learning, the algorithm receives input data without being explicitly told what to do with it. Without human oversight, the programme looks for patterns and structure in the data.

3.2.3 Regression:

Regression is a type of supervised learning job in which the output value is predicted continuously, as opposed to discrete labels as in classification. The process of determining the best fit line (or curve) that minimizes the difference between the expected and actual values is utilized to represent the relationship between a dependent variable and one or more independent variables.

RESULTS AND DISCUSSION:

The dataset we have selected has 100 samples for each crop. A machine learning model can predict more accurately when the training data is in a wide range.

Fig-5.1 Graphical representation of N, P, K, temperature, pH, and rainfall data in the dataset

S No	Machine learning model	Accuracy
1	Decision Tree	0.22
2	AdaBoost Classifier	0.74
3	XGB Classifier	0.76
4	Random Forest	0.76
5	Logistic Regression	0.96
6	SVM Classifier	0.97245
7	KNN Classifier	0.9886

Performance Analysis of Seven Machine Learning Models

4. CONCLUSIONS

The macronutrients content of the soil has been forecasted by the utilization of machine learning techniques. Furthermore, we have examined the macronutrients found in the soil. We have estimated the appropriate crop that can be grown in the specific area based on the analysis. We have trained, verified, and evaluated seven machine learning algorithms for prediction. We obtained superior accuracy in logistic regression, SVM classifier, and KNN classifier out of those seven machine learning techniques.

Because the KNN method has a greater accuracy rate than the others, we have selected it as the efficient algorithm.

We have created a website that accepts inputs in the form of temperature, humidity, pH, rainfall, nitrogen, phosphorus, and potassium. Following analysis, a crop with a higher yield upon cultivation will be shown in the output field. In the future, meteorological information, such as aerial photos of the planet, may be used to feed the models. It is necessary to switch from machine learning models to deep learning models for processing inputs with two or more dimensions. Features may also include geographic information such as longitude, latitude, and altitude (height above sea level).

These studies could be the basis for a mobile application that is built. Features that optimize search history and provide users with general news will enhance the application's user experience. It is possible to create smartphone assistive devices that can require farmers to press just one button.

5. REFERENCES

[1] Anguraj, K., Thiyaneswaran, B., Megashree, G., Shri, J. P., Navya, S., & Jayanthi, J. (2021). Crop recommendation on analyzing soil using machine learning. Turkish Journal of Computer and Mathematics Education, 12(6), 1784-1791.

[2] Senapaty, M. K., Ray, A., & Padhy, N. (2023). IoT-Enabled Soil Nutrient Analysis and Crop Recommendation Model for Precision Agriculture. Computers, 12(3), 61.

[3] Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., … & Saker, L. (2023). Crop Prediction Model Using Machine Learning Algorithms. Applied Sciences, 13(16), 9288.

[4] Farwa, U. E., Rehman, A. U., Khan, S. Q., & Khurram, M. (2020). Prediction of Soil Macronutrients Using Machine Learning Algorithm. International Journal of Computer (IJC), 38(1), 1-14.

[5] Ikram, A., Aslam, W., Aziz, R. H. H., Noor, F., Mallah, G. A., Ikram, S., … & Ullah, I. (2022). Crop Yield Maximization Using an IoT-Based Smart Decision. Journal of Sensors, 2022, 1-15.

[5] Vasu, D., Singh, S. K., Sahu, N., Tiwary, P., Chandran, P., Duraisami, V. P., ... & Kalaiselvi, B. (2017). Assessment of spatial variability of soil properties using geospatial techniques for farm level nutrient management. Soil and Tillage Research, 169, 25-34.

[6] Waiker, V., Pal, M. V., & Gawande, M. A. (2018). Organization learning by analyzing Manufacturing losses-A case study.

[7] Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision