Soil Analysis and Crop Fertility Prediction Using Machine Learning

Adrija Shree[1], Ekansh Singh[2], Ajay Chaudhary[3]

[1,2,3]Student, CSE, Institute of Technology and Management Gida Gorakhpur, Uttar Pradesh, India.

Abstract*: India's economy is built on agriculture. Agribusiness is responsible for 50% of the remaining task in India. Agriculture has a higher level of commitment in the Indian economy than any other division. In any event, farmers used traditional harvesting strategies, resulting in lower yield profitability. Furthermore, soil erosion and its integration is a major contributor to lower yield profitability. This will have an effect on the degree of productivity. Loss of soil nutrients due to various processes is also a motivator to reduce soil richness. Supplements such as potassium (K), nitrogen (N), and phosphorus (P) are essential for plant development. To address these challenges in the agribusiness sector, agricultural advancement is critical, and judicious cultivation is the ideal approach..Crop mutation is a strategy that farmers use after each subsequent crop production. The crop mutation allows the soil to reclaim minerals that were previously used by the crop and use the minerals that were left over to cultivate the new crop. Farmers must experience a decrease in production to determine if the soil has reached the point when it is unfit to yield the specific crop. Accepting a loss in one financial year is critical for a farmer. This research proposes a that would aid in the continuous maintenance of soil fertility. This strategy has been used in many places where a change in crop is made following a loss in yield from constantly cultivating the same crop. When it comes to predicting soil quality, three soil factors are taken into account. Using Machine Learning Techniques, this method provides a solution to the problem indicated above. This research proposes a software-based solution for predicting soil quality using key soil characteristics and variables.*
Key Word*: Machine learning, soil quality, Weka Tool, Artificial Intelligence*

I. INTRODUCTION

India is a country with abundant natural and people resources, and its economy is expanding at a rapid pace. A substantial portion of the Indian economy is based on agriculture, and in order to enhance agricultural practises, it is required to accurately forecast crop output responses, which can be accomplished using Machine Learning. Agricultural soil quality is determined by macro and micronutrient content, such as S, K, pH, C, Mg, P, Ca, B, and others. Our primary goal is to investigate, adapt, and formulate soil features and crop growth variables. The major goal of this research is to look at macro and micro soil parameters that affect crop output, such as organic content and vital plant nutrients, and to use Supervised Learning to determine the rank of a given soil based on previously graded soils. The major goal of this research is to look at macro and micro soil parameters that affect crop output, such as organic content and vital plant nutrients, and to use Supervised Learning to determine the rank of a given soil based on previously graded soils. Using Machine Learning Techniques, this method provides a solution to the problem indicated above. When predicting soil quality, three soil parameters are taken into account:

- Chemical Parameters
- Physical Parameters
- Biological Parameters

Climate warming and other environmental changes have posed a serious danger to agriculture. Machine learning is a critical component of finding practical and successful solutions to this issue. Predicting agricultural yield based on historical data such as weather, soil, rainfall characteristics, and previous crop yield. We used a machine learning technique to accomplish this. We compared the accuracy of numerous machine learning methods, including Random Forest, SVM, Bagging, and the multilayer Perceptron (MLP) Algorithm, which all scored above 90%.In this project, a mobile application was created that forecasts agricultural yields in general and for a specific crop. It also tells the user whether or not it's time to apply fertilizers.

II. LITERATURE SURVEY

Literature Survey is a systematic and thorough search of all types of published literatureas well as other sources including dissertation, these in order to identify as many items as possible that are relevant to a particular topic.

Predicting agricultural products plays a very important role in agriculture. It helps in increasing net produce, better planning and gaining more profits. To achieve better results, westudied a few research papers related to our project topic.

1. Machine learning approach for forecasting crop yield based on climatic parameters
   S.Veenadhari et al. (2014) [13] In this paper, the study was aimed to develop a website for finding out the influence of climatic parameters on crop production in selected districts of Madhya Pradesh

2. Prediction of Crop Yield Using Machine Learning
   Deepak.G et al. (2018), [15] The research of soil and fertilizers is included in this work. The proposed system forecasts the production of a specific crop. It also returns information about the specified fertilizer and makes a recommendation based on the soil attributes and location.

3. Predicting Yield of the Crop Using Machine Learning Algorithm
   P.Priya et al. (2018),[12] This paper uses R programming with Machine Learning techniques. R is the leading tool for statistics, data analysis, and machine learning.

PROBLEM STATEMENT AND SOLUTION

Problem- Farmers nowadays apply fertilizers without considering how their soil is growing, which has an impact on soil health and crop productivity.
As a result, the suicide rate among farmers is rising.
Solution- Assuming this is a problem, we will use machine learning to analysis the nutrients in the soil and determine whether it is fruitful or not.
We will also anticipate the best crop for that soil for that condition, as well as the yield of crops in that location, after assessing the soil nutrients and climate condition of that area.
SYSTEM IMPLEMENTATION: In this project a mobile application has been developed which predicts the crop yield in general and also for a particular crop. Along with that, it also suggests the user if it is the right time to use the fertilizer or not.
We created a model in this project that can determine if the soil is fertile or not, sow crop seed on fertile soil, and finally predict crop production based on various soil characteristics. This system is effective and saves time and cost of the user.
METHODOLOGY: The goal of this study is to look at predictive algorithms that have fewer important features. Data collection and purity are critical for achieving the research work's aims. The methods outlined in each approach, as well as the many resources, are eventually connected up into a single platform to fulfil the goal, as this study is a combination of three approaches. Irrigation data, related meteorological data, fertilizer usage data, and yield statistics were all employed in this study. The information was gathered from various sources and pre-processed. To find the most significant features, the pre-processed data was fed into the most relevant feature selection algorithms. The crop production is predicted using the input dataset's selected attributes as input to predictive algorithms. The procedures adopted for this work are:
1. Data collection
2. Data pre-processing
3. Feature normalization (scaling)
4. Feature selection
5. Applying prediction algorithms (available algorithms and developed hybrid algorithm)
6. Training and testing
7. Cross-validation
8. Performance evaluation

ADVANTAGES
- The system aims to help farmers to cultivate proper crop for better yield production.
- The end user is provided with proper recommendations about fertilizers suitable for every particular crop
- The results of the two algorithms will be compared and the one giving the best and accurate output will be selected.
- Thus, the system will help reduce the difficulties faced by the farmers and stop them from attempting suicides.

III. RESULT

The statistical analysis results of the data used in this research are presented in this segment. Different algorithms had been used to pick the features from the dataset which were more relevant. The significant features were chosen depending on the selection conditions of the feature selection model. It was complicated to find which feature subset gave better predictions. All the available feature subsets had been given in the Random Forest, Bagging, Support Vector machine (SVM), and Multilayer Perceptron (MLP) model to find how they were statistically related between dependent and independent features are the performance matrices Kappa statistic, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative absolute error (RAE) and Root Relative Squared Error (RRSE) of different feature selection models and the matrices calculated by the potential feature subset selected by these

algorithms and table 2 represents detailed accuracy by class for random forest (RF)

IV.CONCLUSION

This strategy is suggested to address the rising prevalence of farmer suicides while also assisting them in becoming financially stronger. The Crop Recommender system assists farmers in predicting crop yields as well as determining which crops to plant. It also informs the user when the fertilizer should be applied.

Using machine learning technologies, appropriate datasets were acquired, analysis and trained. The system keeps track of the user's needs and pulls data from the suggested machine learning model. As a result, the user only needs to input basic information such as soil type and weather.

This is a problem with several classes. In this WEKA tool, we consider Machine Learning (ML) methods including Random Forest, SVM, Bagging, and Multilayer Perceptron with preconfigured sub tools. Kappa statistic, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative absolute error (RAE) and Root Relative Squared Error (RRSE), Accuracy, FPR, Precision, TPR/Recall/Sensitivity, F1- Score, MCC, ROC are some of the performance indicators. These machine learning algorithms' performance varies. In our investigation, we discovered the optimal model to be Random Forest.

References

[1]. Austin, P. C. and Steyerberg, E. W. (2015), 'The number of subjects per variable required in linear regression analyses', Journal of clinical epidemiology 68(6), 627–636.

[2]. Bajardi,P.,Barrat,A.,Natale,F.,Savini,L.andColizza,V.(2011),'Dynamicalpatterns of cattle trade movements', PloS one 6(5), e19869.

[3]. Bajardi, P., Barrat, A., Savini, L. and Colizza, V. (2012), 'Optimizing surveillance for livestock disease spreading through animal movements', Journal of the Royal Society Interface 9(76), 2814–2825.

[4]. Baranyi, J., J´ o´zwiak, A., Varga, L., M´ezes, M., Beczner, J. and Farkas, J. (2013), 'Application potentials of network science, bioinformatics and systems biology to food science', Hungarian Science 174(9), 1094–1102.

[5]. Bell, D. A. and Wang, H. (2000), 'A formalism for relevance and its application in feature subset selection', Machine learning 41(2), 175–195.

[6]. Bennett, J.D.(2016),Value of map sharing between multiple vehicles in the same field while using automated section control, PhD thesis, Kansas State University.

[7]. Berk, J. and Berk, S. (2000), Quality management for the technology sector, Elsevier.