Selecting the Ideal Crop for Specific Soil Types*

1st Md.Shariar Emon Shaikat
dept. of CSE
Brac University
Dhaka, Bangladesh
shariar.emon.shaikat@g.bracu.ac.bd

2nd Syed Shakib Ahmed
dept. of CSE
Brac University
Dhaka, Bangladesh
syed.shakib.ahmed@g.bracu.ac.bd

3rd Ridwanul Hoque
dept. of CSE
Brac University
Dhaka, Bangladesh
ridwanul.hoque@g.bracu.ac.bd

***Abstract*—"Machine intelligence is the last invention that humanity will ever need to make." Nick Bostrom. Using that very invention in order to grow and produce the most demanding material that humans need has made a great impact. Machine learning has been used in various fields and is making huge developments very rapidly. Machine learning can use the agriculture related data in order to yield various crops based on the minerals, level of nutrition, climate and weather factors etc. The goal of this project is to make an algorithm which will predict the suitable crop based on the nutrients of the soil. In order to make the algorithm dataset that are being used contain different minerals, temperature and Ph level of multiple crops. To make this an effective crop suggesting algorithm we used 8 different models such as logistic regression, random forest, naïve bayes, SVM, KNN etc. Through this process we hope that farmers in future do not have to go through the various issue to identify which crop has better growth rate in which soil and increase the production rate and decrease the expenditure**

Index Terms*—Machine Learning, soil, accuracy, agriculture, prediction, nutrition

I. Introduction

Agriculture is Bangladesh's most significant contributor; around 12 percent of the total economy is maintained by the agricultural sector, and Bangladesh has the largest workforce. Despite being the most crucial sector of the economy, it has faced the most challenging problems. To begin with, every year crops are restricted due to arable land, climate changes, rising sea levels, etc. According to the 2019 agricultural census, Bangladesh has been losing cultivable land at a rate of 0.99In addition, this cultivable land needs to be monitored in order to preserve soil fertility. Excessive fertilizer use can ruin soil fertility, planting crops without prior knowledge about soil nutrients and minerals or knowledge of which crop will grow better in which soil. All of these issues can be eliminated with machine learning. To begin with, using machine learning to predict crops and soil based on soil minerals and nutrients will help agriculture avoid all the general and critical problems it faces and increase production, resulting in losses. Our model will predict all these based on the classified data and will learn from the existing data in order to perfect its accuracy and predict much more accurately. Various machine learning models have been used to achieve maximum efficiency in the algorithm. Furthermore, the classified data that are being used in the algorithm contain data on all types of minerals and Ph levels. Moreover, soil prediction using machine learning is a groundbreaking project that deals with the critical aspects of agriculture. The concept of machine learning uses the process of learning to achieve a goal. To achieve a goal, they complete and do a specific process, and through the process, the system gains experience. Gaining those experiences will help the system to make assumptions. These assumptions can used for testing and classification of data, etc. These data are obtained through the system or machine training process. The project aims to revolutionize the approach to soil classification in Bangladesh. There are multiple ways to make this algorithm, but this particular project focused on the minerals and Ph levels of the soil. Taking different parameters into consideration, we want to make the algorithm as accurate as possible to ensure reliability, which, in the near future, can be used by Bangladesh to develop the country's backbone of the economy.

II. Problem Statement

In order to provide more precise guidance for crop selections, our project seeks to employ machine learning methods. We will perform a more thorough investigation by directly analyzing the raw soil data. The leading aim is to create a complicated multi-layered system where the exact amount of all essential nutrients, such as potassium, phosphorus, and nitrogen, as well as all the environment's parameters, such as temperature, humidity, pH, and rainfall patterns, play a role. This approach comes into play by studying the characteristics that are unique to the different types of soils. Hence, it can correctly classify the various soils and crops suitable to be grown under each soil. In our effort, we aim to help them improve their yields by using safe farming methods and applying the knowledge domain with innovative machine learning algorithms to give farmers valuable data on their production.

III. Literature Review

The actual idea of the method is to make a group of various soil types according to their characteristics and then predict which plants will succeed in each category using machine learning. To "preprocess" data means to get the data ready for testing. Encoding, normalization, statistical analysis, and maintenance are all part of this. The features are selected with the aim of a greater understanding of the data. As a result, the model gets better while the number of dimensions becomes lower. Many models are applied, including Decision Trees,

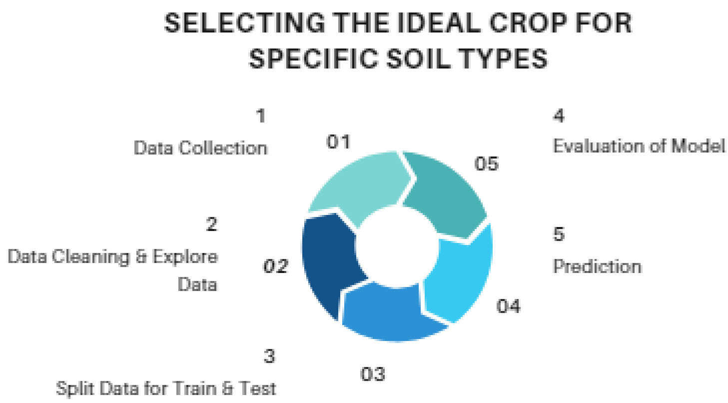

Fig. 1. Figure of soil types

XGBoost, Naive Bayes, Random Forest, and Kth Near Neighbors. To get the most out of each approach, we used a number of parameter-tuning techniques, including grid and random search algorithms. The models are thoroughly examined and changed based on how well they classify large groups of crops, using a refined set of features to make sure that predictions are accurate and reliable. This method not only improves the accuracy of crop predictions but also helps farmers make intelligent decisions by giving them specific crop suggestions based on complete data analysis. This makes intelligent farms more effective places to grow crops. Using the Internet of Things (IoT) along with advanced machine learning methods is a big step ahead in farming that aims to meet the needs of modern growing ecosystems.

This comprehensive review of recent research shows how different ML methods can be used to improve crop yields by predicting soil nutrients as well as quality levels. To calculate Soil Quality Indices (SQIs) from the physical and chemical characteristics of soils and assess soil elements like Organic Carbon, Phosphorus, and Potassium, methods like Neural Networks, Decision Trees, Vector Machines, and Random Forests have been used. According to these studies, multi-component analyses, among those that use group models and Principal Component Analysis (PCA), are more accurate than single-predictor models. Soil fertility classification and nutrient recommendation optimization have also made use of techniques like clustering using K-means and Fuzzy Logic. The successful use of ML to boost the accuracy of soil tests and later adaptation of agricultural practices to unique soil conditions is highlighted by the better results of ensemble and hybrid methods that combine different machine learning models to identify different soil properties.

IV. Methodology

A. Data Preprocessing

- Data Cleaning: We meticulously address any missing values within the dataset.Employing advanced imputation techniques, where necessary. We ensured the integrity and completeness of our dataset.
- Exploratory Data Analysis(EDA): A comprehensive analysis of the distribution of soil attributes was conducted. Through advanced visualization techniques, we gained profound insights into the intricate relationships between various soil properties and crop types.
- Encoding: Here we changed the class column; categorical variables, such as crop categories, were meticulously converted into numerical representations. Here we used the One-Hot-Encoding technique to transform qualitative data into a format suitable for quantitative analysis, thereby empowering our model to effectively interpret and learn from these features.

B. Feature Selection

We identified vital factors like soil pH, moisture content, temperature, and nutrient levels crucial for soil classification and crop recommendation. Our chosen features include nitrogen, potassium, phosphorus, rainfall, and pH, known for their significant impact on soil fertility and crop growth. By prioritizing these essential variables, we aim to develop precise models that effectively capture soil-crop dynamics for informed agricultural decisions.

C. Model Architecture

In our project, we used a mix of machine-learning methods to help classify soil types and recommend suitable crops. Here's a breakdown of what we did:

- Logistic Regression: We used logistic regression to solve our multi-class crop categorization challenge. This technique readily expands to handle several classes, which makes it appropriate for our goal. Using the ”one-vs-rest” technique, we trained distinct logistic regression models for each crop class. Cross-validation was performed to evaluate model performance, and measures including accuracy, precision, recall, and F1 score were used. This method allowed us to create a reliable system for outlining crops based on soil characteristics and environmental considerations.
- Random Forest:For our multiclass crop recommendation challenge, we used random forest, a collaborative ensemble of decision trees. Each decision tree gave information into the best crop option based on soil characteristics and environmental considerations. We combined the trees' projections and chose the most popular crop as our proposal. This technique enabled us to create a comprehensive crop recommendation solution that ensures accurate forecasts based on a wide range of soil conditions and environmental variables.
- Gradient Boosting: It operates by learning from mistakes and increasing accuracy with each model it creates. We trained a sequence of models, each of which corrected the faults of the preceding one, to forecast the optimal crop for various soil and environmental circumstances. Cross-

validation ensured that our suggestions were accurate and dependable.

- Decision Tree: In our study, decision trees were used to make multiclass crop recommendations. They operate as step-by-step guides, making judgments depending on soil and environmental conditions. Each phase evaluates a single component, such as nitrogen levels, and divides the data accordingly. This process is repeated until a leaf node makes a final choice on which crop is best. Decision trees provided clear and precise advice for our agricultural demands.
- Naive Bayes: This approach relies on probabilities to categorize data points, with the assumption that each characteristic contributes independently to the classification. It's like making a guess based on how probable each component is to result in a specific soil type or crop idea.
- K-Nearest Neighbors (KNN): We went to KNN, which is comparable to asking around for guidance from neighbors. To determine where a new data point belongs, it looks at the nearby data points. It's similar to asking the people next door what they think based on their past observations.We went to KNN, which is comparable to asking around for guidance from neighbors. To determine where a new data point belongs, it looks at the nearby data points. It's similar to asking the people next door what they think based on their past observations.
- Support Vector Machine (SVM): SVM's ability to categorize crops accurately was crucial to the success of our endeavor. SVM achieved accurate multiclass classification by determining the most significant separation line between various crop groups based on soil and environmental variables, providing important insights for agricultural decision-making.

V. RESULT

In the world of precision agriculture, choosing the best crop for a given soil type is critical for increasing output and sustainability. To accomplish this, multiple machine learning methods were used and evaluated for their accuracy in predicting crop suitability. Our study looks at the performance of seven prominent classifiers: Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machines (SVM), Naive Bayes, Decision Tree, and K-Nearest Neighbors (KNN). Naive Bayes achieved the highest accuracy of 99.55 percent, followed by Random Forest at 99.32 percent and Decision Tree at 98.64 percent. SVM achieved an impressive 96.14 percent accuracy, while Logistic Regression and KNN achieved 94.55 percent and 97.05 percent, respectively. Surprisingly, Gradient Boosting had the lowest accuracy at 9.818 percent, indicating potential limitations in its application to this job. These findings highlight the effectiveness of ensemble approaches such as Random Forest and Naive Bayes for making precise crop recommendations based on soil characteristics, as well as the importance of selecting algorithms that are adapted to the complexity of agricultural datasets. Such insights can considerably impact agricultural decision-making processes, leading to increased crop output and resource utilization efficiency.

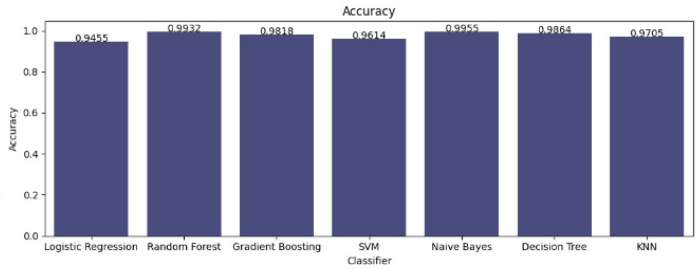

Fig. 2. Figure of soil types

VI. FUTURE ENHANCEMENT

The proposed model that has been selected for predicting the soil type based on the minerals and suggesting the appropriate crop has been tested through multiple model testing like KVM, random forest, logistic regression, and 4 more algorithms, we have achieved the maximum accuracy. Although this project has been made based only on minerals and Ph level of the existing data it can be done through several other data like weather and climate change and many other factors that play a vital role in agriculture. Preliminary we are using existing data but to enhance the algorithm we can gather real time data while using it and it can give much more accurate predictions. Moreover, in future to get more accuracy we can use hyper parameter tuning to achieve the maximum accuracy. Furthermore, the project is compatible with sensors or any type of devices like mobile, or different modern machinaries, so in future this algorithm can gather data using those technologies and make predictions which will increase the accuracy percentage as well. Lastly,to keep the model relevant to agriculture it is very important to understand the limitations and make necessary adjustments as the project advances.

VII. ACKNOWLEDGMENT

Throughout the course of this project's development, our honorable faculty, Dibyo Fabian Dofadar, provided priceless advice and unfaltering support, and we are eternally grateful to him. He was helpful in our topic selection, and his helpful feedback led us to improve our claims and achieve greater understanding of what was covered.

The mentorship had a significant impact on how we addressed the project, and we are truly humble of the experience and knowledge that he provided. This project would never have been approved of had he not been so encouraging and supportive. I want to thank our supportive faculty once more for all the help he has provided us.

VIII. CONCLUSION

Machine learning is a game-changing tool in the ever-changing field of problem-solving, not only that it also demonstrates how these advanced technologies are helping humankind in solving various critical problems in different sectors. Machine learning has continuously made lives and work easy for the very fact it has the capabilities of learning on its own and uses that experience to do work much better than the previous. Like in any other sector machine learning is playing a great role in the agriculture sector, from the various trained data machine learning has successfully made a lot of predictions which helped in the development of a lot in the sector. This project has successfully achieved its goal of predicting the best soil for the crops based on the data that include Ph level and minerals and nutrients. The project has successfully achieved the maximum accuracy through the model we trained containing multiple different models. In the future, the project will be able to gain more data and will be able to predict the accurate outcome successfully with those data.

REFERENCES

[1] Rani, S., Mishra, A. K., Kataria, A., Mallik, S., Qin, H. (2023b). Machine learning-based optimal crop selection system in smart agriculture. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-42356-y

[2] Ijraset. (n.d.-b). Soil classification and crop suggestion using machine learning. IJRASET. https://www.ijraset.com/research-paper/soil-classification-and-crop-suggestion-using-ml

[3] Elbaşi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A. Y., Saker, L. (2023). Crop prediction model using machine learning algorithms. Applied Sciences, 13(16), 9288. https://doi.org/10.3390/app13169288

[4] Dey, B., Ferdous, J., Ahmed, R. (2024). Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables. Heliyon, 10(3), e25112. https://doi.org/10.1016/j.heliyon.2024.e25112

[5] JK, P. (2020). Soil Classification and Crop Suggestion using Machine learning techniques. ResearchGate.

[6] Suruliandi, A., Mariammal, G., Raja, S. P. (2021). Crop prediction based on soil and environmental characteristics using feature selection techniques. Mathematical and Computer Modelling of Dynamical Systems, 27(1), 117–140. https://doi.org/10.1080/13873954.2021.1882505

[7] Gosai, D., Raval, C., Nayak, R. J., Jayswal, H. D., Patel, A. (2021). Crop Recommendation System using Machine Learning. https://www.semanticscholar.org/paper/Crop-Recommendation-System-using-Machine-Learning-Gosai-Raval/4780fa47758b038ace07b72caff7e393a44f04d8

[8] Awais, M., Naqvi, S. M. Z. A., Zhang, H., Li, L., Zhang, W., Awwad, F. A., Ismail, E. a. A., Khan, M. I., Raghavan, V., Hu, J. (2023). AI and machine learning for soil analysis: an assessment of sustainable agricultural practices. Bioresources and Bioprocessing, 10(1). https://doi.org/10.1186/s40643-023-00710-y