ISSN No:-2456-2165

# Agrilyst: The Crop Advisor

Sonal Chaudhari
Assistant Professor,
Dept. of Computer Engineering,
Datta Meghe College of Engineering,
Airoli, Navi Mumbai, India

Shaikh Haroon Shahadatali, Pandey Govind Parashuram, Vadalia Dhruvin Dharmesh
Bachelor of Engineering, Dept. of Computer Engineering,
Datta Meghe College of Engineering,
Airoli, Navi Mumbai, India

**Abstract: As India is an agrarian country, its economy depends mostly on the growth of agricultural yields and agro-industrial products. Data mining is an emerging research area in crop yield analysis. Yield prediction is a very important issue in agriculture. Every farmer is interested in how much yield he can expect. Discuss the various related attributes such as location, pH from which soil alkalinity is determined. In addition, the percentage of nutrients such as nitrogen (N), phosphorus (P) and potassium (K). Nutritional value of the soil in this region, you can determine the amount of precipitation in the region, the composition of the soil. All these data attributes will be analyzed, they will train the data with various suitable machine learning algorithms to build a model. The system comes with a model to predict crop yield precisely and accurately, and gives the end user the appropriate recommendations on the required fertilizer ratio based on the soil and atmospheric parameters of the land, which they improve to increase the yield and the income of the raise farmers.**

***Keywords:-*** *Machine Learning, Crop prediction, Decision tree, Random Forest, Fertilizer recommendation, Heroku, Crop recommendation;*

## I. INTRODUCTION

INDIA is a densely populated country and weather conditions change randomly to secure the world's food resources. Framers face serious problems during drought. Soil type plays an important role in crop yield. Suggesting fertilizer use can help growers make the best decision for their growing situation. The number of studies on information and communication technologies (ICT) can be used for crop yield prediction. By using data mining we can also predict the yield of the crop. By fully analyzing the above data, we can suggest a better crop for a better yield to the farmer. In order to get a better yield, we must take into account the type of soil and soil fertility, as well as one of the main factors, rainfall and the availability of groundwater. If it is dry land, it is better to go for commercial crops, and if it is wetland, it is better to go for wheat and sugarcane. There are 15 agroclimatic regions in India, these regions are divided according to a type of country. Each agroclimatic region can produce specific crops. Based on this, we must suggest to the farmer which crop is the best among the crops of these climatic regions. Achieving the maximum harvest with the minimum yield is the ultimate goal of the project. Identifying problems early and addressing them can help farmers improve crop yields. Predicting crop yields is important research that contributes to food security. For a better understanding of crop yield, we need to examine the huge amount of data using a machine learning algorithm in so that it gives the exact yield for that crop and suggests a better crop to the farmer. Improving harvest yield is the central goal of precision farming and means better understanding harvesting using information technology methods. The primary goal of precision farming is profitability and sustainability. Agriculture has always been the backbone of our country. Today the climatic conditions fluctuate very often. So it is difficult to grow and understand crops weather conditions. We need to use some technology to find or understand the details of crop and guide farmers to manage accordingly, and besides, fertilizer is also one of the main factors to manage accordingly. If more or less fertilizer is used in the field, the soil may lose its fertility and the crop may not bring the expected yield. So fertilizer also becomes the main factor. First of all, understanding the temperature conditions is very important for India as we can improve Indian economy with the help of crop forecast as it plays an important role in Indian economy. In general, machine learning algorithms will predict the most efficient power output. Yield used to be predicted based on farmers' previous experiences, but now weather conditions can change drastically, leaving them unable to estimate yield. Therefore, the technology can help them to predict the yield of the crop, whether they choose that crop or not. The machine learning model will understand the harvest pattern and yield based on various conditions and predict the yield of the area to be cultivated. Also, they apply fertilizers in random amounts because they don't know how much nutrients the plants need. Because of this, it directly affects the crop yield, and also causes soil acidification and further damage to the top layer. To solve this problem, we have developed a system that uses machine learning algorithms to improve farmers. Our system suggests the best possible harvest for a given piece of land based on the soil content and various other natural parameters such as temperature, rainfall, etc. In addition, the system provides all the necessary information about the soil content required for a specific crop, and consequently farmers would be recommended the amount of fertilizer required for the crop. Therefore, by using our system, farmers can grow a new variety of crop under the appropriate conditions, reducing the rate of soil contamination. All of this would increase

ISSN No:-2456-2165

your profit margins and ultimately improve your quality of life and improve the financial health of our country.

## II. LITERATURE SURVEY

[A] Ashwani kumarKushwaha[1] describes crop yield predictionmethods and a suggest suitable crop so that it will improve theprofit for the farmer and quality of the agriculture sector. Inthis paper for crop yield prediction they obtain large volumedata, it's been called as big data (soil and weather data) usingHadoop platform and agro algorithm. Hence based repositorydata will predict the suitability crop for particular conditionand improvement crop quality.

[B]. Chawla, I. et al. (2019, August) used fuzzy logic for crop yield prediction through statistical time series models. They considered parameters like rainfall and temperature for prediction.Their prediction was classification with levels 'good yield' , 'very good yield'.

[C] Girish L [3] describe the crop yield and rain fall prediction using a machine learning method. In this paper they gone through a different machine learning approaches for the prediction of rainfall and crop yield and also mention the efficiency of a different machine learning algorithm like liner regression, SVM, KNN method and decision tree. In that algorithm they conclude that SVM have the highest efficiency for rainfall prediction.

[D]. Armstrong, L. J. et al.(2016, July) used ANNs for the prediction of rice yield in the districts of Maharashtra, India. They considered climatic factors namely (considering range) temperature, precipitation and reference cropevapo transpiration. The records were collected from Indian Government repository from 1998 to2002 [7]. Tripathy, A. K. et al. (2016, July) were same authors who used support vector machines to predict the rice crop yield with same features as the previous paper mentioned.

[E]. Manjula, A et al.built crop selection and to predict the yield which considered various indexes like vegetation, temperature and normalized difference vegetation as factors. They distinguished between climate factors and agronomic factors and other disturbances caused in the prediction for better understanding .

[F]TamilNadu, India.They have considered actors like soil, temperature, sunshine, rainfall, fertilizer, paddy, and type of pest used and other factors like pollution and season.

## III. REQUIREMENT ANALYSIS

After looking at the problems farmers face, such as B. Lack of knowledge about what crops to grow and what fertilizers to use for a particular crop, by considering each and every aspect of farming, as well as natural parameters, we came to the conclusion to solve the above problems. issue efficiently. We concluded that there is a need to develop a system that will help farmers in our country get all the answers to their questions about what crops to grow and what fertilizer to use. This can be done by implementing various machine learning strategies and algorithms. A detailed dataset would be used to train the model, containing all the key parameters needed to make accurate predictions. For better usability and user experience, a web application would be developed that can also be accessed from anywhere, which is an added bonus. Ultimately, all of these would help farmers predict the right crops to grow and the right fertilizers to use, and prevent farmers from making wrong decisions, thereby improving farmers' financial situation..

## IV. PROJECT DESIGN

### *A. DESIGN GOALS*

Design goals consist of different designs that we will implement in our system. Harvest prediction and fertilizer recommendation through machine learning. The main goal of the project is to predict the perfect harvest for the country in question and also the optimal amount of fertilizer for the country. This system is made with different designs like Use Case Diagram, Component Diagram, Activity Diagram, Implementation Diagram, System Architecture Diagram. After creating these different charts and based on these charts we will develop our project. Here are the things this system will do.harvest forecast. fertilizer recommendation. Fruit Prediction – The system takes user input based on various predefined parameters and displays the most suitable fruit for each crop. Fertilizer Recommendation – Based on various input parameters, the system also has the ability to recommend the amount of fertilizer required for the soil.

### *B. SYSTEM ARCHITECTURE*

Agrilyst using machine learning will predict the crop variety and fertilizer recommendation for the user based on various symptoms and the information provided by the user such as temperature, humidity, NPK value and many more general information about the tickets. The Agrilyst system architecture, using machine learning, consists of multiple datasets through which we compare and predict user inputs, then the datasets are converted into smaller datasets and sorted from there based on the sorting algorithms. Later, the classified data is processed in the machine learning technologies through which the data is processed and entered into the Agrilyst model using all of the above user inputs. In general, after the user enters the above information, the processed data is combined and compared in the system prediction model, and finally predicts the crop and fertilizer. An architecture diagram is a graphical representation of a set of concepts that are part of an architecture, including its principles, elements, and

components. The diagram explains the system software in the perception of the system overview.

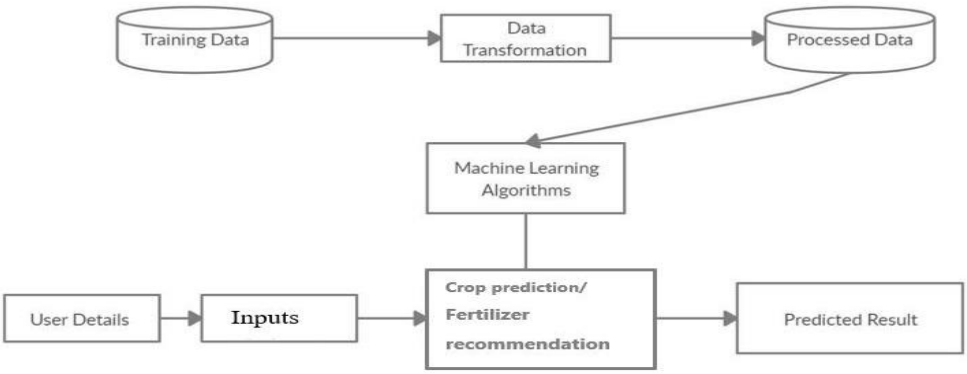



Fig. 1: SYSTEM ARCHITECTURE

*C. COMPONENT DIAGRAM*

A component diagram, also called a UML component diagram, describes the organization and wiring of a system's physical components. Component diagrams are often created to model implementation details and to verify that all aspects of the system's required function are covered by the planned development. Here, the component diagram consists of all the main components sed to build a system. So the design, the algorithm, the file system and the datasets are all linked together. The datasets are used to compare the results and the algorithm is used to process these results and give the correct precision and the design UI is used to display the result in the, So all the components are linked together.

*D. DEPLOYMENT DIAGRAM*

A deployment diagram shows the configuration of the compute nodes at runtime and the components on them. Implementation diagrams are a type of structure diagram used to model the physical aspects of an object-oriented system. Here the deployment diagram shows the final phase of the project t and also shows what the model looks like after all the processes have been run and deployed to the machine. Starting from the way the system processes the information entered by the user and then compares this information using data sets, it trains and tests this data using algorithms such as Decision Tree, Naive Bayes, Random Forest. Then the system processes all this data and information and outputs the desired result on the interface.

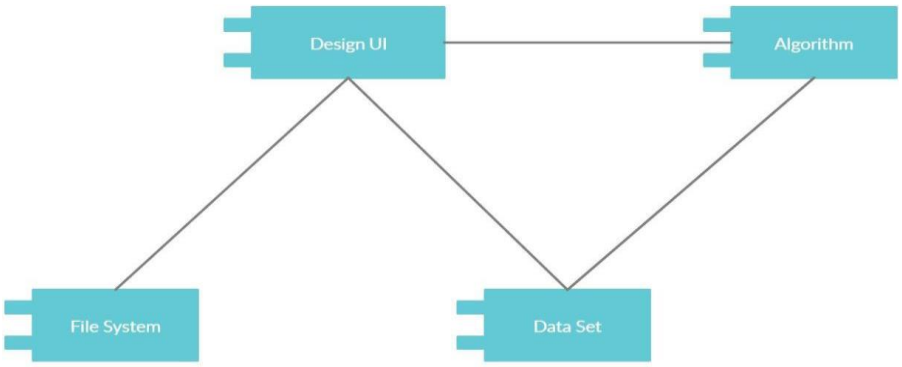



Fig. 2: COMPONENT DIAGRAM

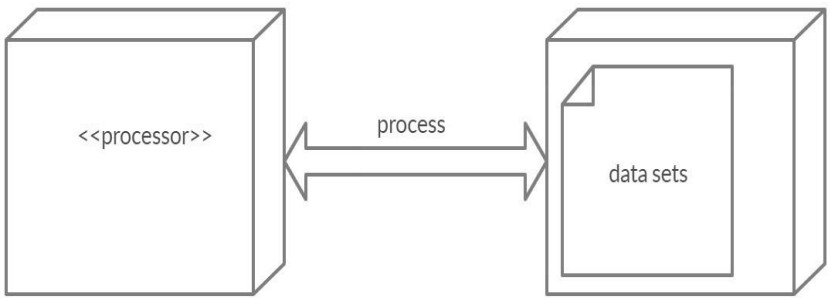



Fig. 3: DEPLOYMENT DIAGRAM

ISSN No:-2456-2165

## V. MPLEMENTATION

Agrilyst, the crop advisor, is a project in which we combine two of the most important aspects of the agricultural sector, namely the crop to choose and the optimal amount of fertilizer needed for each country. In the crop prediction system, the user (farmer) can make various inputs such as soil data, precipitation, temperature, location, etc. to predict the best possible harvests for your land. In the fertilizer recommendation system, the user (farmer) can enter the soil data and the type of crop he is growing into the system, so that the system recommends fertilizer (if necessary) to improve crop yield. Two main aspects of farming i.i.e. crops and fertilizers are combined in one system/project, which means that a fertilizer recommendation is added alongside the crop forecast, giving farmers answers to all their questions in a single place, such as: B. "What crop do you want to grow here?" (using harvest forecast) and "Which of all fertilizers can be used for this particular crop?" (Using the fertilizer recommendation). A detailed dataset is used to provide the best possible prediction and recommendation considering all aspects of farming such as: B. Soil content (NPK values), temperature, pH, precipitation, moisture, condition, etc.to achieve the best possible results. It uses numerous supervised machine learning algorithms like Decision Tree, SVM, Naive Bayes, Random Forest, Logistic Regression and some newer algorithms like XGBoost and finally on the endpoint the algorithm that gives us the best possible accuracy for the selected dataset would be selected for better prediction and

- generation can also handle a large number of information variables without eliminating variables.
- 8. Training and Experimentation on Datasets The Agrilyst model is trained on the dataset to accurately make the prediction and build the confusion matrix.In this project more than 3 different algorithms are compared:
  - • Decision Tree Algorithm
  - • Random Forest Algorithm
  - • Naive Bayes Algorithm

recommendation. Finally, for the GUI part, a web application is developed that can be accessed by any user (farmer) from anywhere and at any time, and at the same time becomes interactive for users. The step-by-step implementation is given below:

- The user enters the system and is given the option of crop prediction or fertilizer recommendation.2. If the user selects crop forecast, he will be redirected to the crop forecast page where he will be given several input parameters to complete.
- Based on the provided inputs and the dataset, the most optimal algorithm is used to grow the right crop.
- When the user selects fertilizer recommendations, they are taken to the fertilizer recommendations page where they are given several input parameters to complete.
- Based on the inputs provided and the dataset, the most optimal algorithm is used to obtain the optimal amount of fertilizer needed.
- Data Collection and Dataset Preparation This involves gathering information from various sources such as government records, weather history, micronutrient requirements of plants and then pre-processing the dataset, which removes all unnecessary data and extracts important features from the data.7. Developing a probabilistic model and deep learning (RNN) approach forAgrilyst In this step, you will develop a probabilistic model and deep learning approach based on RNN that can be run effectively on large agricultural databases. And decision tree
  - • SVM
  - • XGBoost
  - • Logistic Regression .

Implementation and analysis in a real scenario The trained and tested predictive model is implemented in a real scenario created by human experts and used to further improve the methodology.

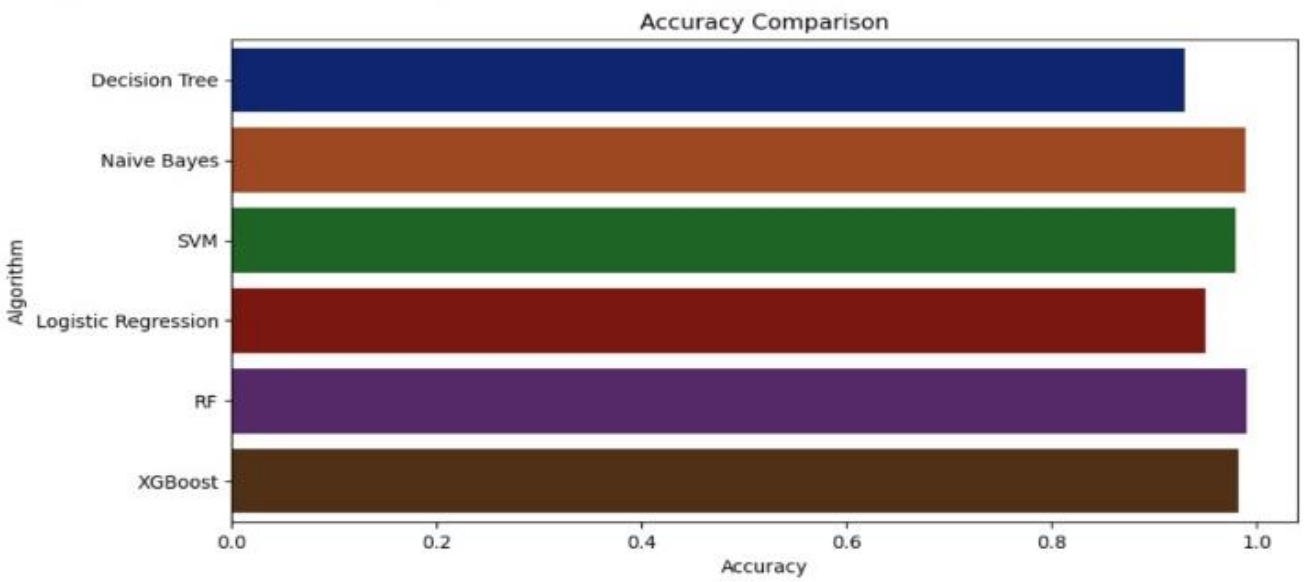



Fig. 4: Accuracy of algorithm

ISSN No:-2456-2165

### *A. DECISION TREE ALGORITHM*

Decision tree induction is decision tree learning from training tuples with class labels. A decision tree is a tree structure similar to a flowchart. Decision tree induction is a nonparametric approach to building classification models. Finding an optimal decision tree is an NP-complete problem. The techniques developed to build decision trees are computationally cheap, allowing models to be built even when the size of the training set is very large. Decision trees, especially smaller trees, are relatively easy to interpret. The decision tree provides an expressive representation for learning discrete-valued functions. The decision tree algorithms are fairly robust to the presence of noise, especially when it comes to methods to avoid overfitting

### *B. RANDOM FOREST ALGORITHM*

It is an ensemble classifier using many decision tree models; can be used for both regression and classification. The results can provide information of varying importance and accuracy. A random forest is the classifier consisting of a collection of k-tree structured classifiers, where k are independent and identically distributed random trees and each random tree consists of the voting unit for the input classification. Random Forest uses the Gini index for classification and for determining the final class
in each tree. The final class of each tree is aggregated and voted on the weighted values to create the final classifier. The operation of the random forest consists in choosing a random seed that randomly extracts a collection of samples from the training data sets while preserving the class distribution.

### *C. SUPPORT VECTOR MACHINE*

Support Vector Machine or SVM is one of the most popular supervised learning algorithms used for both classification and regression problems. However, it is mainly used for machine learning classification problems.

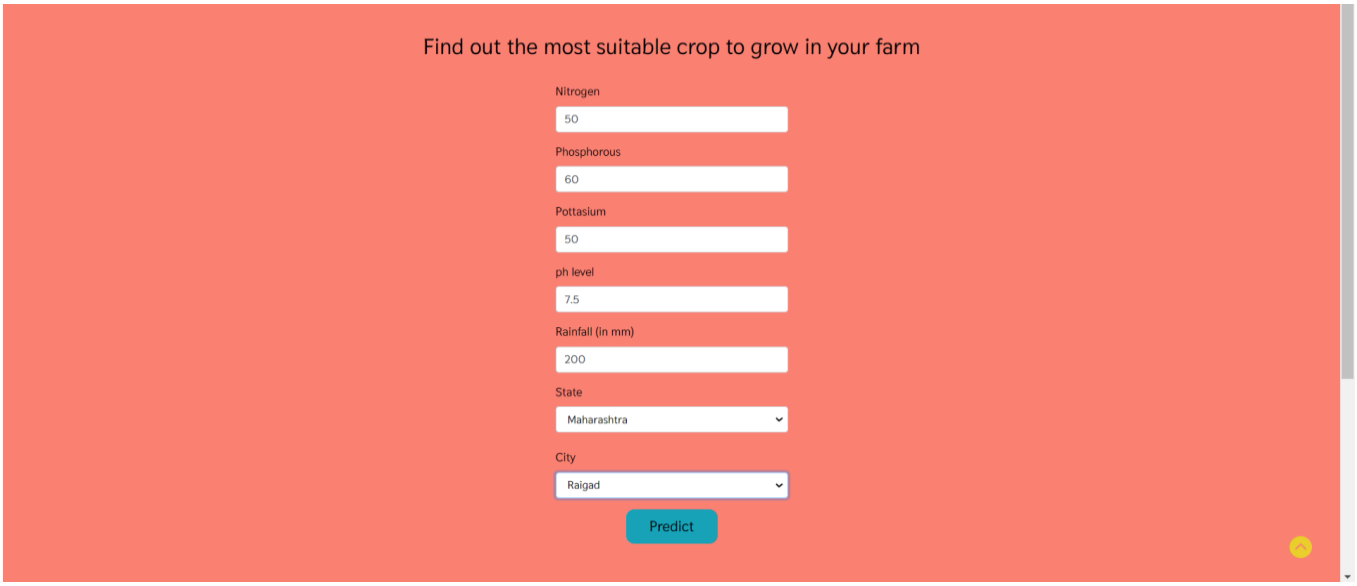
The goal of the SVM algorithm is to create the best decision line or boundary that can partition the n-dimensional space into classes so that we can easily place the new data point in the correct category in the future. This best decision boundary is called the hyperplane. SVM selects the extreme points/vectors that help create the hyperplane. These extreme cases are called support vectors and hence the algorithm is called a support vector machine.

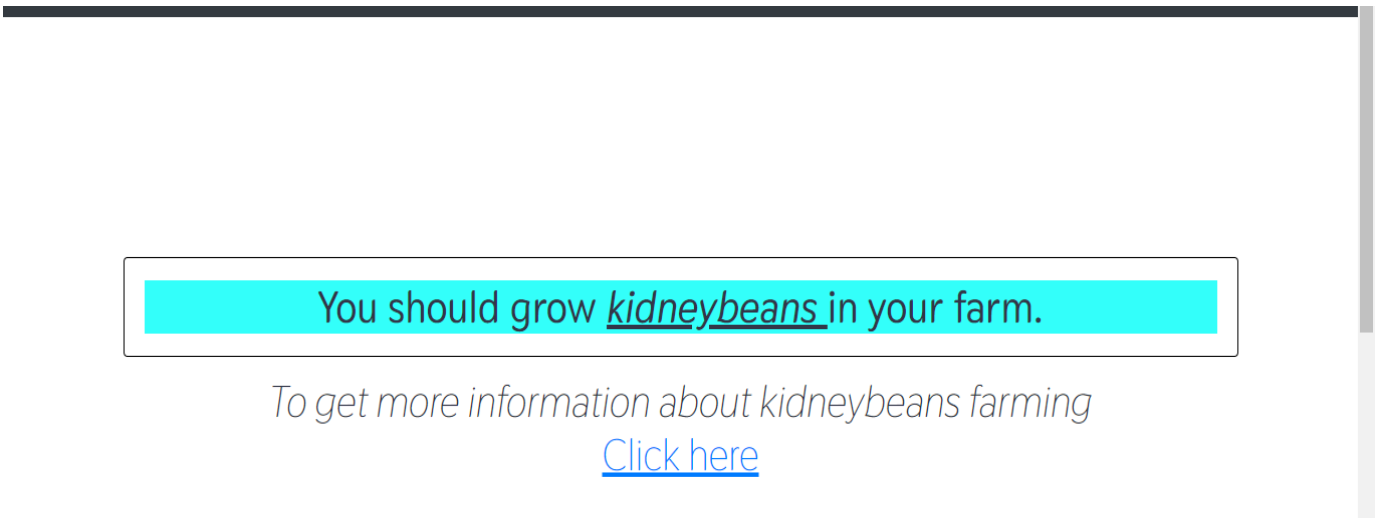
## VI. TECHNOLOGIES USED

- Programming Language :- Python
- ML Libraries Used :- NumPy ,Pandas ,Sklearn ,etc.
- Algorithms Used:- Decision tree, Random Forest, SVM, etc.
- Software Used :- Microsoft Visual Studio Code(Jupyter notebooks)
- User interface :- Web Application using FLASK
- Hardware Used :- Intel Core Processor, 8 GB RAM, 2-5 GB HDD, Integrated Graphics, LAN/Wi-Fi
- O.S. Used :- Windows OS 10/11 (64 bit)

## VII. RESULTS

ISSN No:-2456-2165




The K value of your soil is high.

Please consider the following suggestions:

| | |
|---|---|
| 1 | *Loosen the soil* deeply with a shovel, and water thoroughly to dissolve water-soluble potassium. Allow the soil to fully dry, and repeat digging and watering the soil two or three more times. |
| 2 | *Sift through the soil*, and remove as many rocks as possible, using a soil sifter. Minerals occurring in rocks such as mica and feldspar slowly release potassium into the soil slowly through weathering. |
| 3 | Stop applying potassium-rich commercial fertilizer. |
| 4 | Mix crushed eggshells, crushed seashells, wood ash or soft rock phosphate to the soil to add calcium. Mix in up to 10 percent of organic compost to help amend and balance the soil. |
| 5 | Use NPK fertilizers with low K levels and organic fertilizers since they have low NPK values. |
| 6 | Grow a cover crop of legumes that will fix nitrogen in the soil. This practice will meet the soil's needs for nitrogen without increasing phosphorus or potassium. |

## VIII. CONCLUSION AND FUTURE SCOPE

Farmers generally use the hit-and-miss method, which results in a waste of land and resources, or even overgrowth of crops. Here we try to break down all those tax barriers by giving you an accurate and justifiable model built by machine learning to identify the right crop to grow on your farms and the fertilizers for your respective crops. This would help them quantitatively and qualitatively improve the production of their crops and maintain the quality and nutrient content of the soil for later reuse. All of this would ultimately help farmers improve their income and quality of life.

By classifying soil according to soil type, soil type and the macronutrients nitrogen (N), phosphorus (P) and potassium (K) present in the soil, along with soil temperature, pH and electrical conductivity, the agricultural operator is provided with the suggested the most suitable crop along with the right fertilizer to enrich the soil and increase productivity.A yield forecast for the crop desired by the farmer is also provided. Alternative crops that can be grown for the given growing season and that are also suitable for the level of NPK contained in the soil are considered to make suggestions to the farmer

Regarding the future evaluation if farmers grow a particular crop they could face some pre-harvest problems or diseases in crop. In this case you can upload the photos of the harvest and the soil report. The AI model can then identify problems and offer likely solutions to them. The data set may be updated in the future based on weather conditions and new harvest data may be added.Live weather forecasts also help users predict crop water needs and help farmers reduce crop damage due to rain or drought.

## REFERENCES

[1.] Ashwani kumar Kushwaha, Swetabhattachrya “crop yield prediction using agro algorithm in hadoop”.
[2.] Bang, S., Bishnoi, R., Chauhan, A. S., Dixit, A. K., & Chawla, I. (2019,August). Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA,SARIMA,andARMAXmodels.
[3.] Girish L, Gangadhar S, Bharath T R, Balaji K S, Abhishek K T “CropYield and Rainfall Prediction in Tumakuru District using Machine Learning”.
[4.] Gandhi, N., Petkar, O., & Armstrong, L. J. (2016, July). Rice crop yield prediction using artificial neural networks. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp.105-110).IEEE.
[5.] Manjula, A., & Narsimha, G. (2015, January). XCYPF: A flexible and extensible framework for agricultural Crop Yield Prediction. In 2015IEEE 9th International Conference on Intelligent Systems and Control(ISCO)(pp.1-5).IEEE.
[6.] Mariappan, A. K., & Das, J. A. B. (2017, April). A paradigm for riceyield prediction in Tamilnadu. In 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development(TIAR)(pp.18-21).IEEE.