

A Novel Framework for Effective Utilization of Agriculture Land With Yield Prediction Using Random Forest Classifier

¹G. Sri Charani, ²M. Krishna Satya Varma

¹M. Tech Scholar, ²Assistant Professor, Department of IT, S.R.K.R Engineering College, Bhimavaram, India. ¹sricharani186@gmail.com, ²varamantena@srkrec.ac.in

Abstract: Manmade resource or procedure intelligence is branch of study of applied science, that is incorporating all types of human applications. albeit our farmers area unit well knowledgeable with their productivity and coming up with, we have a tendency to imagined to offer further info to get a lot of profit with following parameters. during this situation, we've got to divide our analysis add 2 completely different parameters like soil properties and roots of plants. state, beginning a agriculture state, its economy dominantly depends upon gardening yield development and agro business things. info Mining is Associate in Nursing arising analysis field in crop yield examination. Yield forecast could be a considerable issue in farming. Any granger is keen on knowing what proportion yield he's getting to anticipate. Examine the various connected properties like space, hydrogen ion concentration esteem from that pH scale of the dirt is resolved. aboard it, level of supplements like atomic number 7 (N), phosphoric (P), and atomic number 19 (K) space is employed aboard the utilization of outsider applications like arthropod genus for climate and temperature, quite soil, supplement price of the dirt around there, live of precipitation around there, soil arrangement are often resolved. For Soil analysis, we've got to urge the options from the land and water perceptive capability. Next, choose plants like Rice, Sugarcane and its varieties. 2 completely different options area unit extracted and mixed with single propagation rule. we have a tendency to planned, Random Forest classifier rule for getting the optimum result from the agriculture sectors. supervised learning rule of Random Forest. Associate in Nursing ensemble of call trees builds the "forest", with the "bagging" methodology. The material methodology is a mixture of learning models will increase the general result.

Keywords — Random Forest Algorithm, Machine learning, Agriculture analysis

I. INTRODUCTION

Human beings are fond of necessary agriculture because it forms the basis for food security. Farming plays a vital to people since it frames the reason for food correct security. It assists with developing the harvests of food and raise the creatures with agreement to natural variables. India's farming is made out of numerous yields, with the preeminent yields. Indian ranchers likewise develop natural products like oilseed sugarcane etc. More than 70% of the provincial families rely upon horticulture. Farming contributes about 17% to the all-out Gross domestic product and gives work to more than 60% of the populace. Agribusiness is specific significant incomes creating areas of India and a wellspring of endurance. Different occasional, financial and natural components impact the yield creation however capricious substitutes in these elements lead to an extraordinary misfortune to rrs. These dangers can be measured when suitable numerical or factual procedures are applied on information identified with soil,

climate and past yield. With the appearance of information mining, crop yield can be anticipated by getting helpful bits of knowledge from this horticultural information that guides ranchers to choose the harvest they could need to plant for the approaching year prompting most extreme profit.

The horticultural perception incites ranchers to change their agrarian creation and shipment. What's more, it contributes the public authority's stockpile strategy and value adjustment measures. To work together in the country, like agribusiness perceptions, there have been a ton of examination and framework building. In any case, due to an Earth-wide temperature boost and unusual climate, farming gauge is getting more mind boggling and more troublesome. We can perceive the weightiness the existing circumstance from the cabbage wave rehashed each year. Farming creation of one year from now is influenced from crop level of going before year, value, utilization designs, and imported agrarian items. It mulls over the conclusion of a blend of various components. The expectation of

horticultural items called divine indent. so as to support reliable prediction service, performed several studies and government funded domestic many governments connected institutes. To predict the quality, determined statistical data. Thanks to lack of enumerate power process technology of information storage, several knowledges weren't stacked. Large amount of technologies and distributed storage increased power computing surged. According to recent advances, bigdata technology as recently spotting, improved greatly.

II. APPLICATIONS OF MACHINE LEARNING(USED)

The applications of machine learning embrace email filtering, detection of network intruders or malicious insiders operating towards an information breach, optical character recognition (OCR), image recognition, speech recognition, executing to rank, pc vision, drug discovery and personalized treatment in healthcare industry, fraud detection in finance industry, product recommendations and improved customer services in retail industry, dynamic pricing and sentiment analysis in travel industry, search engine result refining, videos surveillance and internet media services. [8]

A. Python programming

Python is associate taken, vital level and loosely helpful artificial language. Python is unceasingly created and junk assembled. It maintains various programming norms, together with coordinated (particularly, procedural), object-arranged programming. Python uses vigorous creating and a mixture out of reference checking and a cycle-perceiving garbage knowledgeable for memory the board. It what is more incorporates dynamic name objective (late confining), that ties technique and variables throughout program execution. Python licenses designers to explain their own types mistreatment categories, that are generally used for object-arranged programming. New events of categories are created by job the category (for example, Spam Class () or Eggs Class ()), and also the categories are case in point of the meta category sort. Python's large normal library, systematically mentioned as perhaps the simplest strength, provides instruments work associate extreme range of tasks. For net facing applications, numerous normal arrangements and shows, for example, Copy and hypertext motion protocol are maintained. It consolidates modules for making graphical UIs, interfacing with social knowledge bases, making pseudorandom numbers, conniving with self-emphatic preciseness decimals, dominant typical verbalizations, and unit testing.

- Most Python executions (tallying CPython) consolidate a read-eval-print circle (REPL), allowing them to operate as asking line middle person that the client enters decrees increasingly and moves results in general time.
- Different shells, together with Latent and IPython, add any limits like improved auto-realization, meeting state support and punctuation highlight.
- Python's improvement is coordinated normally with the Python language design Suggestion (Kick) live, the basic framework for putting forward major new options, gathering neighborhood on problems and chronicling Python arrange choices. [] Python secret writing vogue is roofed in Energy eight. shocking PEPs square measure investigated and commented on the Python socio-economic class and therefore the coordinating consultatory cluster.
- Improvement of the programming language contrasts and progression of the C-Python backing execution. The listing python-dev is that the elementary spoken language for the language's new development. Unequivocal problems square measure inspected within the event bug huntsman worked with at bugs.python.org. Progression initially happened on a self-worked with source-code store running unsteady, till Python affected to GitHub issue in Jan 2017.
- CPython's public conveyances are available 3 types, perceived by that piece of the structure variety is augmented: Backward-contrary forms, wherever code is relied upon to interrupt and will be physically ported. The initial segment of the variant variety is raised. These deliveries happen inconsistently—for instance, variant 3.0 was delivered eight years after 2.0.
- Major or "highlight" discharges, regarding at regular intervals, square measure to a good extent viable nevertheless gift new highlights. The second piece of the rendition variety is increased. every vital adaptation is upheld by fixing issues for quite long whereas once its delivery.
- Bugfix discharges, that gift no new highlights, happen regarding like mechanism Associate in Nursing square measure created once an adequate variety of bugs are fastened crucial since the last delivery. Security weaknesses square measure in addition fastened in these deliveries. The 3rd and last piece of the shape variety is raised.

III. SCOPE OF THE PROJECT

The extent of the undertaking is to decide the harvest yield of a space by considering dataset for certain highlights which are significant or identified with crop creation like temperature, dampness, precipitation, and creation of the harvest in earlier years. To anticipate a consistent worth, relapse models and characterization are utilized. It is a regulated procedure. The fundamental concentration here is to lower the expense work by tracking down the best fit-line. The yield work works with to the ranchers to manage

the plant subtleties for accomplishing the normal Yield expectation utilizing Irregular Woodland Classifier. We utilize the black box and white box testing boundaries for effective yield.

IV. EXISTING SYSTEM

This approach is delineated by a mud info base gathered from the ranch, crop given by business specialists, accomplishment of boundaries, as a sample, through soil testing workplace dataset. The knowledge from soil testing workplace dataset given to proposal framework can utilize the knowledge and do cluster model with larger half casting a ballot methodology utilizing support vector machine and ANN as students to recommend the yield for a website specific boundary with high exactitude and effectiveness.

A. Limitations

- Less accuracy and data dependency
- Mismatching about yield and Monsoon situations
- Less number of neurons are utilized in ANN application

V. PLANNED SYSTEM

The total framework is planned utilizing Python. Diverse dataset like harvest, crop yields dataset, Area, soil and harvest supplements, manure dataset is assembled from different source like agrarian books etc., horticultural sites. Sporadic forest area is a coordinated man-made intelligence estimation reliant upon bunch learning. Outfit learning may be quite acknowledging wherever you be a part of various data types of estimations or the same computation on various events to outline an even more great assumption model. The self-assertive woodlands estimation joins distinctive computation of a comparative kind. Discretionary Forest computation can be mostly used for gathering and backslide issues.

a. Advantages

- The arbitrary woodland area unit various trees and every tree is ready on a set of various data.
- Random woodland calculation is steady if another information purpose is conferred within the dataset the overall calculation is not influenced.
- Accuracy and knowledge normalization applied.

VI. SYSTEM IMPLEMENTATION

A. Architecture

In our venture center around AI which incorporates administered learning models. Design is one which distinguishes the Applied model in various constructions and different perspectives on the framework. Fig.4.1 shows building plan of proposed arrangement of the undertaking. Above design plainly clarifies about how the segments of the planned framework impart among themselves beginning from preprocessing of information. This proposed structure

can discover the harvest yield. This model gives clear image of colossal measure of information catch and before processing of information to eliminate the undesirable information, for example, Invalid and later introduced in it. During preprocessing step, we spit the dataset into preparing and testing dataset. Train dataset to pick out the harvest yield available in the dataset utilizing fitting managed learning calculations. Apply the AI procedures which are used for discovering crop yield for any of new information happened in the information. After this information obtaining reasonable AI calculation should be put in to register proficiency and ability of the presented model, here we have applied arbitrary woodland classifier.

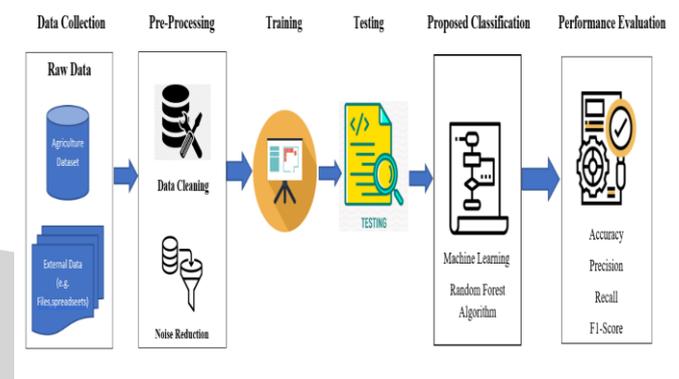


Fig -1: System Architecture

B. Use case

A use case diagram could be a graphical depiction of a user's potential interactions with a system. A use case diagram shows numerous use cases and differing types of users the system has and can typically be in the midst of different forms of diagrams furthermore.

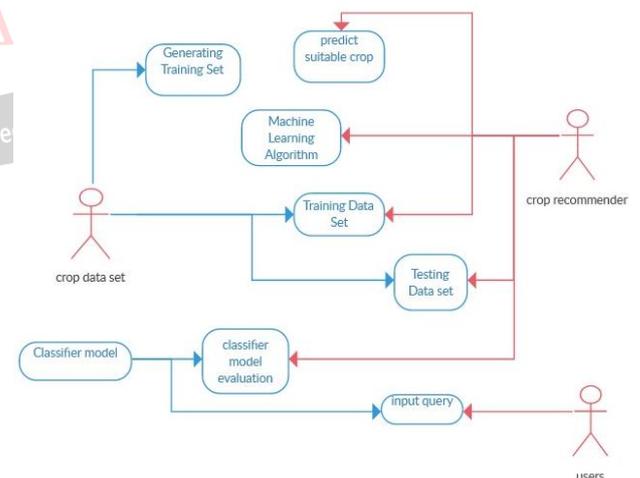


Fig -2: Use Case Diagram

C. Sequence diagram

A sequence diagram may be a sort of interaction diagram as a result of it describes how—and in what order—a cluster of objects works along. These diagrams area unit utilized by software system developers and business professionals to grasp

necessities for a brand-new system or to document an existing method.

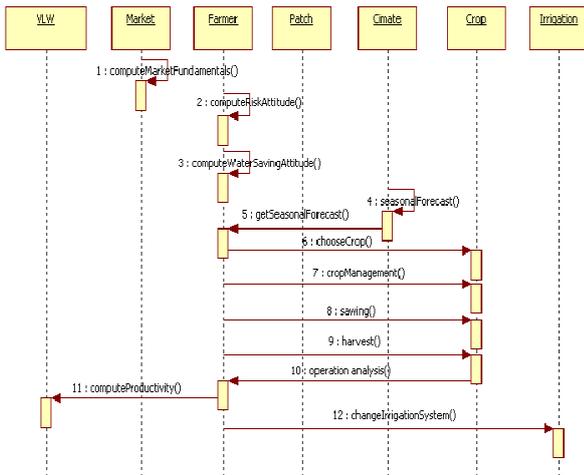


Fig -3: Sequence Diagram

D. Flow diagram of data

A data flowchart (DFD) maps out the flow of knowledge for any method or system. It uses outlined symbols like rectangles, circles and arrows, and short text labels, to indicate knowledge inputs, outputs, storage points and therefore the routes between every destination.

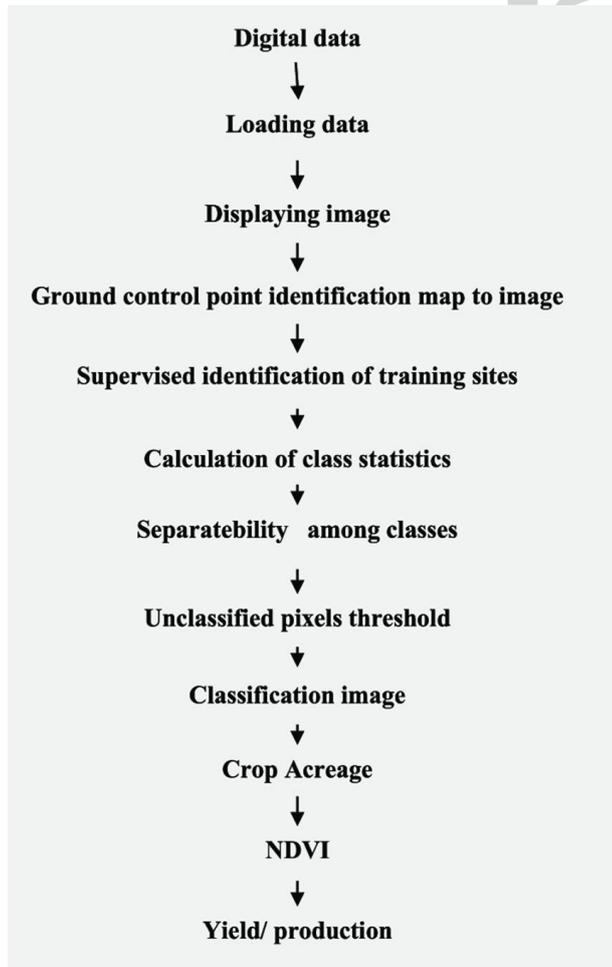


Fig -4: Data flow diagram

VII. METHODOLOGY

In our project, we used four modules for obtaining the result of crop yield prediction

1. Dataset Maneuvering
2. Data Preprocessing
3. Training the dataset
4. Mapping with plants & accuracy
5. Algorithm Implementation

A. Dataset maneuvering

This dataset was worked to give appraisals of the quantity of food created by homesteads of various sizes across the state. We developed this dataset by blending rural censuses and broadly delegate family test reviews that straightforwardly estimated crop creation and additionally editing area by ranch size. Our dataset catches 51.1% of worldwide harvest creation and 52.9% of worldwide cropland region (i.e., arable land and lasting yield region as revealed in the Food and Farming Association's measurable data set. We ran into a few methodological issues while fitting the hidden information expected to develop this dataset.

region	region_score	soil_value	result
west godavari	1	500	50000
Anatpur	2	400	45000
godavari	4	900	90000
kaveri	5	1500	150000
krishna	7	600	60000
mahnadi	8	200	20000
cooium	9	700	72000
kuruvadweep	10	100	10000

Table -1: Region wise agriculture – Classification

B. Data Preprocessing Flow(flow)

Information pre-handling an information mining procedure which is used to substitute the crude details in a helpful and productive structure. In this model, dataset is pre-processed to fill the missing qualities, the fitting data run and isolating the handiness.

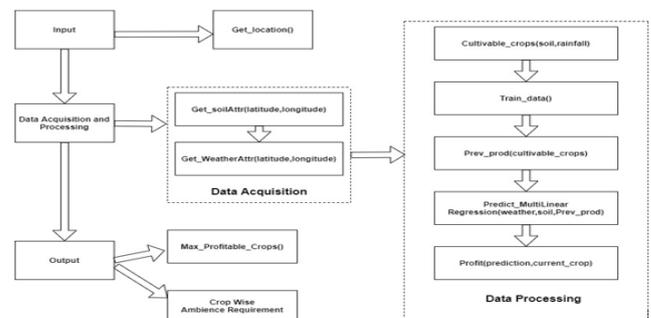


Fig -5: System Design

The numerous information and missing parts can have data cleaning. To genuinely deal with it includes treatment of

missing information, information cleaning is finished and so on.

- **Missing Data:** The present data emerges when some information is absent in the information. It is very well may be dealt with differently.

1. Ignore the tuples:

This technique is suitable once the dataset we've a really vast and totally different qualities are absent within a tuple.

2. Fill the Missing values in the list:

There are varied ways in which to modify do that endeavor. you'll be able to opt to fill the lost characteristics genuinely, by quality mean or the foremost probable price.

- **Noisy gathered Data:** Boisterous data is an insignificant data that can't be deciphered by machines. It is extremely well may be created because of broken data assortment, information section blunders and so on.

1. Data Reduction:

Since data mining is a strategy that is planned to deal with enormous measures of information. By working with enormous volume of information, investigation caught more enthusiastically in such various cases. To dispose this, we use data decrease procedure. It expects to build the lessen information and capacity proficiency stockpiling and examination costs.

C. Training the dataset

This progression incorporates preparing and testing of obtained data. The stacked data segregated into two sets, for example, planning and testing of information. Preparing sets is planned with the preparation sets and during the preparation stage information is to be tested in the wake of gaining from past perceptions. The last information is framed and is prepared by AI module. Highlight extraction should smooth out the plan of information needed to address a gigantic dataset. Its will likely concentrate helpful qualities from information. The attributes incorporate with the high, low and mean temperature, air stickiness, soil pH, precipitation. This progression incorporates preparing and trail out of the information. These stacked data is confined into two sets, for example, getting ready and trail out of information. Preparing sets is planned with the preparation sets and during the preparation stage of information is to test in the wake of gaining from past perceptions. The last information is framed and is prepared by AI module. These formulae for calculating parameters are:

- Load Harvest Dataset
- Data Pre-handling
- Feature extraction
- Split Information for information investigation



Fig -6: Preprocessed dataset

D. Mapping with plants & accuracy

There are 2 strategies for yield planning: Customary techniques and far off detecting techniques. Traditional techniques are frequently convoluted, exorbitant, tedious and they can't be run in huge scope. In this manner it is key to utilize less expensive/quicker techniques for crop yield assessment. Distant detecting information has the budding and the potential to give spatial data at worldwide scale; of highlights and wonders on earth on a practically ongoing premise.

- New data and advanced communication latest technologies makes the field-level crops management best operational and easy to achieve for our country farmers.
- Agricultural equipment uses a technology of Precision Agriculture.
- Positioning system
- Geographic latest information systems

Accuracy Horticulture enables ranchers to utilize crop inputs all the more viably including composts, pesticides, till age and water system water. More powerful utilization of data sources implies more prominent harvest yields and qualities, without contaminating the climate. At present season of expanding costs of input, diminishing ware costs and ecological concerns, ranchers and group of Government specialists is searching for better approaches to build effectiveness, slice expenses and buy in to supportable farming. Regardless of the disservices to accuracy cultivating, there is extraordinary expectation that using this innovation will enormously improve the use of ranchers who choose to get on this innovation.

E. Algorithm Implementation

• Random Forest Classifiers

Irregular timberlands area unit AN outfit learning calculation that may use for arrangement that's Anticipating an absolute reaction variable and that they

can further be used for relapse which has foreseeing a continuous reaction variable. Irregular backwoods relapse and characterization models work a gathering of selections tree models to a bunch of knowledge. for every tree, the data area unit recursively compound to additional homogenized units, that is unremarkably alluded to as hubs, to increase the consistency of the reaction variable. Split focus depends on upsides of indicator factors. Consequently, factors accustomed half the data area unit viewed as important logical factors. Irregular woodlands work separates selections of trees to a predefined variety of bootstrapped informational indexes. The Anticipated price of an all-out reaction may be a methodology of the categories from all the individual fitted selections of trees, and therefore the anticipated price of a standardized response is that the mean fitted reaction from each one of the independent trees that occurred due to every bootstrapped check. The capricious forest calculation will rank the overall significance of each indicator variable.

Sporadic forest may be a classifier that advances from call trees. It really involves numerous call trees. To portray another event, every call tree offers an inspiration to enter data; unpredictable woodlands accumulate the orders and picks the foremost given a balloting kind figure a job because the result.

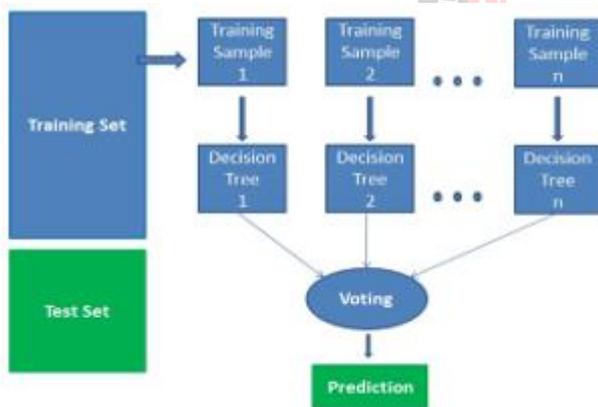


Fig -7: Random Forest – Prediction

Moreover, a set of highlights is at random chosen from the discretionary highlights to grow the tree at each hub. every tree is full-grown while not pruning. Basically, random forest permits an outsized variety of weak or weakly-correlated classifiers to make a robust classifier [9].

• Algorithm Steps

Step 1: Region_Location, Region_score, Soil_value, Weather datasets initialized.

Step 2: Data preprocessed is performed.

Step 3: In Training the soil pH values, temperature and rainfall conditions will be trained.

Step 4: Random Forest Classifier Algorithm is applied.

Step 5: Based up on the given inputs the specific crop will be suggested to the farmer.

Step 6: Guidelines such as accuracy, recall and precision are shown.

VIII. EXECUTION

```
import numpy as np
from flask import Flask, request, jsonify, render_template
import pickle
app = Flask(__name__)
model = pickle.load(open('model.pkl', 'rb'))
@app.route('/')
def home():
    return render_template('index.html')
@app.route('/predict',methods=['POST'])
def predict():
    """ For rendering results on HTML GUI"""
    int_features = [int(x) for x in request.form.values()]
    final_features = [np.array(int_features)]
    prediction = model.predict(final_features)
    output = round(prediction[0], 2)
    message = ""
    price = int(float(format(output)))
    if price <= 10000:
        message = "kuruvadweep"
    elif price > 10000 and price <= 20000 :
        message = "mahnadi"
    elif price > 20000 and price <= 35000:
        message = "black rice"
    elif price > 35000 and price <= 50000:
        message = "kuruvai"
    elif price > 50000 and price <= 70000:
        message = "basmathi"
    elif price > 70000 and price <= 90000:
        message = "jasmine"
    elif price > 90000: message = "parboiled rice"
    return render_template('index.html', prediction_text='You
    should choose rice plant - '+ message)
    # return render_template('index.html', prediction_text='You
    should choose rice plant ' + message +' with estimated
    Amount Rs. {}'.format(output))
@app.route('/predict_api',methods=['POST'])
def predict_api():
    """For direct API calls trough request"""
    data = request.get_json(force=True)
    prediction = model.predict([np.array(list(data.values()))])
    output = prediction[0]
    return jsonify(output)
if __name__ == "__main__":
    app.run(debug=True)
# Importing the libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import pickle
```

```
dataset = pd.read_csv('agri.csv')
dataset['region'].fillna(0, inplace=True)
dataset['soil_value'].fillna(dataset['soil_value'].mean(),
inplace=True)
X = dataset.iloc[:, :3]
#Converting words to integer values
def convert_to_int(word):
word_dict= {'godavari':4, 'kaveri':5, 'krishna':7, 'mahnadi':8,
'cooum':9, 'kuruvadweep':10, 'seven':7, 'eight':8, 'nine':9,
'ten':10, 'eleven':11, 'twelve':12, 'zero':0, 0: 0}
return word_dict[word]
X['region']= X['region'].apply(lambda x : convert_to_int(x))
y = dataset.iloc[:, -1]
#Splitting Training and Test Set
#Since we have a very small dataset, we will train our
model with all available data.
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
#Fitting model with training data
regressor.fit(X, y)
# Saving model to disk
pickle.dump(regressor, open('model.pkl','wb'))
# Loading model to compare the results
model = pickle.load(open('model.pkl','rb'))
print(model.predict([[6, 9, 100]]))
```

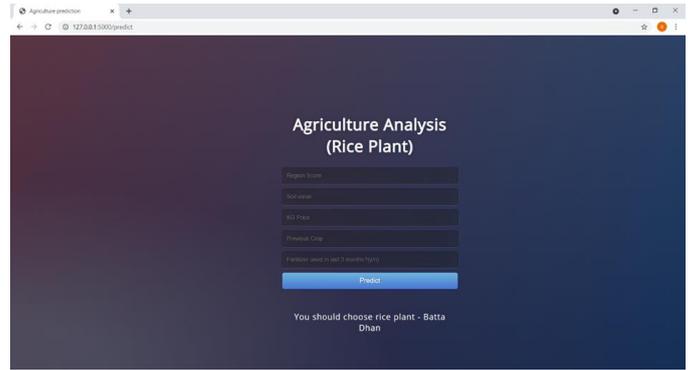


Fig -10: Result screen

IX. RESULTS

An overall soil investigate archive will commonly offer qualities which incorporates pH [12], rate regular matter (OM), phosphorus, and considerable potassium; in any case, different measurements might be phenomenally valuable for a superior data of the dirt's capacity to keep up and supply nutrients, just as the dirt fruitfulness steadiness. Fig shows the example soil test document.

Sample	Ph	P lbs/ac	K lbs/ac	OM %	Ca lbs/ac	Mg lbs/ac	CEC	Micro Nutrients (lbs/ac)					
								S	Zn	B	Fe	Mn	Cu
1	6.5	48	416	3.2	6.111	735	20.33	16	1.3	1.5	120	15	1.7
2	6.2	78	472	3.4	5.376	777	23.02	17	0.7	1	147	12	1.8
3	6.3	64	388	3.2	5.229	1134	20.46	20	0.7	0.9	102	20	1.6
4	6.3	76	508	3.4	5.733	1155	26	25	0.7	1	135	13	2.1
5	5.8	62	462	3.4	5.062	1029	24.57	15	0.9	0.8	117	15	1.5
6	6.2	84	464	3.7	6.615	756	22.49	14	0.8	0.9	141	16	1.9
7	6.1	62	496	3.7	6.216	1071	26.5	16	1.1	0.8	133	18	1.7
8	6.2	52	424	3.6	6.258	750	21.4	14	1.2	0.9	98.5	12	1.5
9	6.3	72	572	3.9	6.972	861	21.66	17	0.9	1	117	13	2.1
10	6.1	44	392	3.2	5.964	903	23.2	13	0.8	1.2	108	11	1.9

Table -1: Sample Soil Test Report

Cation Trade Limit (CEC) and Base Immersion (BS) are crucial qualities on a dirt test report. For legitimate richness plan these realities are generally significant. CEC is usually surely known by utilizing numerous ranchers because of the reality it's far emphatically related with soil surface (grouping of sand, sediment, and mud). Besides, the sum and kind of dirt, along with percent OM, gently affect CEC. Soils with over-the-top CEC (~20+) ordinarily have better water-holding capacity, mud substance, OM, and supplement bestowing energy. Soils with low CEC (<12) normally require numerous ideal supplement programs over the span of the creating season for ideal harvest creation, especially for supplements which incorporates nitrogen and potassium. Utilizing the CEC from the supplement extraction methodology, you'll have the option to figure the BS. BS addresses the section of the CEC. Soils with a reasonable richness have the resulting appropriate Base Immersion stages:

- Calcium – 65% -75%
- Magnesium – 10% -15%
- Potassium – 3% - 6%

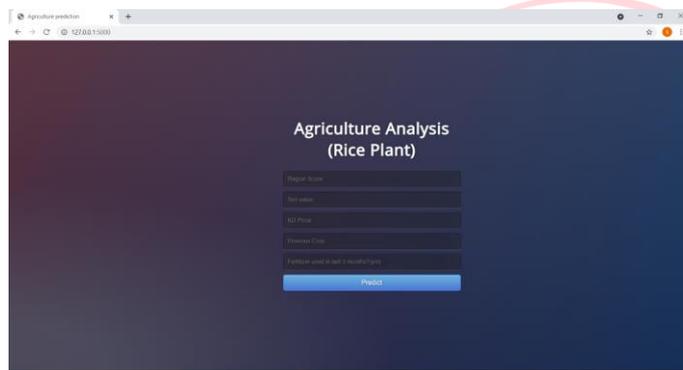


Fig -8: value entering screen

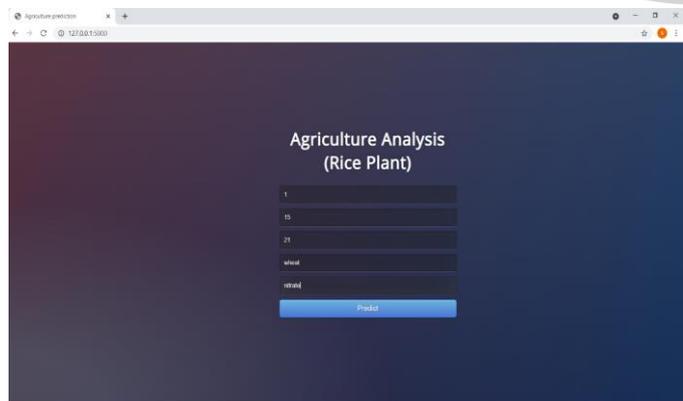


Fig -9: Values entered screen

- Hydrogen - much lower than 10%
- Sodium - much lower than 1%

X. CONCLUSION

This venture is tried utilizing AI and assesses the exhibition by utilizing Irregular solid ground calculation. The accomplishment is higher than framework would facilitate in higher developments of the agriculture acts of our nation. additional it alright is also accustomed reduce the misfortune looked by the ranchers and upgrade the harvest respect improve capital in farming. The models is improved by coordinative this with totally different divisions like cultivation, sericulture, and others towards the farming advancement of our nation. The proposed framework thinks about the data identified with soil, climate and past year creation and recommends which are the best productive yields that can be developed in the pertinent natural condition. As the developed framework rattles off every single imaginable yield, it assists the rancher in vigorous of which harvest to expand. Likewise, this developed framework considered over the previous creation of information which will assist the rancher with getting mastery into the curiosity and the charges of various yields in market. As greatest kinds of yields will be covered under this framework, rancher may become acquainted with about the reaping which may never been evolved.

REFERENCES

- [1] R. Kumar, M. P. Singh, P. Kumar and J. P. Singh, "Crop Selection Method to maximize crop yield rate using machine learning technique," 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2015, pp. 138-145, doi: 10.1109/ICSTM.2015.7225403.
- [2] Ramesh, D., and B. Vishnu Vardhan. "Analysis of crop yield prediction using data mining techniques." International Journal of research in engineering and technology 4, no. 1 (2015): 47-473.
- [3] Zhong, Huaiyang, Xiaocheng Li, David Lobell, Stefano Ermon, and Margaret L. Brandeau. "Hierarchical modeling of seed variety yields and decision making for future planting plans." Environment Systems and Decisions 38, no. 4 (2018): 458-470. DOI <https://doi.org/10.1007/s10669-018-9695-4>
- [4] Bondre DA, Mahagaonkar S. Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms. International Journal of Engineering Applied Sciences and Technology. 2019;4(5):371-6.
- [5] Bhanumathi, S., M. Vineeth, and N. Rohit. "Crop Yield Prediction and Efficient use of Fertilizers." In 2019 International Conference on Communication and Signal Processing (ICCSP), pp. 0769-0773. IEEE, 2019.
- [6] Dr. Manjunath M, Venkatesha G, & Dinesh S. (2020). Agricultural Crop Yield Prediction and Efficient Use of Fertilizer Using Machine Learning. Journal of Android, IOS Development and Testing, 5(1), 1–13. <http://doi.org/10.5281/zenodo.3693626>.
- [7] Anusorn Charleonnann, Thipwan Fufaung, Tippawan Niyomwong, Wandee Chokchueyattanakit, Sathit Suwannawach, Nitat Ninchawee (2016), "predictive analytics for chronic kidney disease using machine learning techniques", The 2016 Management and Innovation Technology International Conference
- [8] Kajaree Das, Rabi Narayan Behera (2017), "A Survey on Machine Learning: Concept, Algorithms and Applications", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 2.
- [9] <https://www.datacamp.com/community/tutorials/random-forests-classifier-python>
- [10] <https://blog.jcharistech.com/2020/09/23/how-to-split-dataset-into-training-and-testing-dataset-for-machine-learning/>