



AMRITA
VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY

Amrita School of Engineering
Chennai – 601 103, Tamil Nadu, India.

A Report Submitted to
AMRITA SCHOOL OF ENGINEERING
Inpartial fulfilment of the Requirements for the
Degree of Bachelor of Technology in
COMPUTER SCIENCE & ENGINEERING

SUBJECT : Foundation of DataScience(19CSE304)

Submitted to: Dr.Sountharjan
[Assistant Professor
CSE Department]

Submitted by :

GAURI SANTHOSH--[CH.EN.U4CSE20023]
GOKUL SIDDARTH S--[CH.EN.U4CSE20024]
GOLLU TWIKIL--[CH.EN.U4CSE20025]
GUHAN N K--[CH.EN.U4CSE20026]
GUNDE ELIJAJER – [CH.EN.U4CSE20027]
HARDEE KRISHNA P P--[CH.EN.U4CSE20028]
HRUSHIKESAVA REDDY –[CH.EN.U4CSE20029]

Build a predictive model to recommend the most suitable crops to grow in a particular farm based on various parameters using crop recommendation Dataset.

ABSTRACT:-

- ❖ Agriculture is first and foremost factor which is important for survival. Machine learning (ML) could be a crucial perspective for acquiring real-world and operative solution for crop yield issue. Considering the present system including manual counting, climate smart pest management and satellite imagery, the result obtained arent really accurate. This paper focuses mainly on predicting the yield of the crop by applying various machine learning techniques. The classifier models used here include Logistic Regression, Nayee Bayes and Random Forest, out of which the Random Forest provides maximum accuracy.
- ❖ The prediction made by machine learning algorithms will help the farmers to come to a decision which crop to grow to induce the most yield by considering factors like temperature, rainfall, area, etc. This bridges the gap between technology and agriculture sector.

INTRODUCTION :-

- ✚ Agriculture crop recommendation is a new generation bubble that is engaging the masses. In most of the cases, the farmers are not

well aware of the kind of crops they should be growing in their farms.

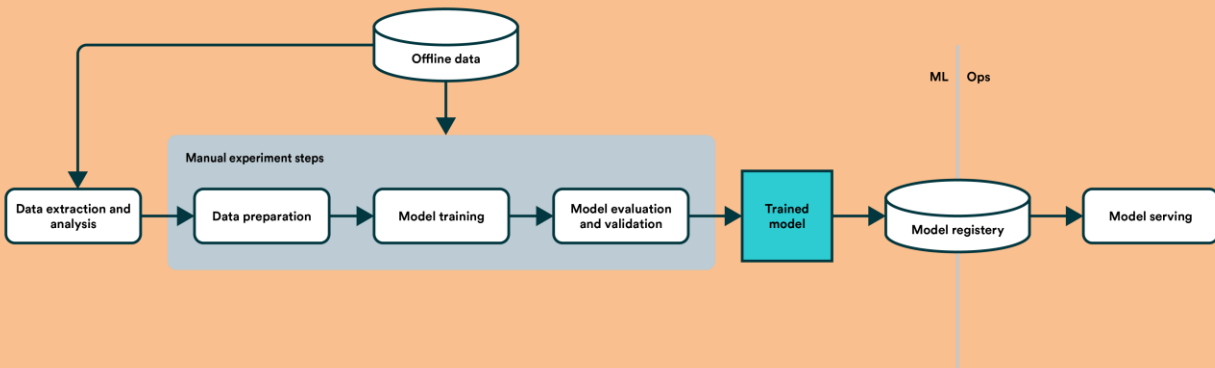
- ✚ This leads to a lot of confusion and affects the productivity. This is why we are focusing on figuring out the best crop to grow in order to get optimum yield.
- ✚ We have gathered a dataset built by augmenting datasets of rainfall, climate and fertilizer data available for India. This will give us a better idea of the trends of crops considering different environmental and geographical factors.
- ✚ We can use this dataset to create a machine learning model for predicting the best suitable crop to grow at a particular place. Machine learning can prove to be the turning point of the agriculture industry.
- ✚ By predicting the right crop to be grown, we will help the farmers to decide the raw materials and other resources required much earlier than they would have figured it out otherwise.
- ✚ This will eradicate the problem of nutrients deficiency in fields occurring because of planting wrong crops which can scale down the production efficiency in a compound manner.
- ✚ India is still lacking behind in finding technological solutions for agriculture which is the primary source of income for about 50% people in the country. Promoting more scientific solutions is the need of the hour in order to take the agriculture industry of India to greater heights.
- ✚ The main idea of the model is to provide the farmers with an ideal recommendation for growing crops taking into consideration the factors such as composition of soil, the environmental factors like temperature, humidity, rainfall and the geographical influence.



DESIGN AND METHODOLOGY :

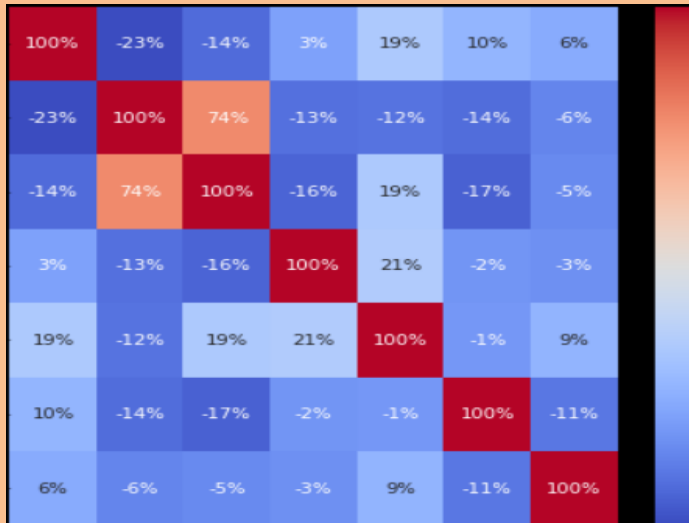
- ❖ The main ideology of the research revolves around the concept of identifying the most suitable crop to be grown with the help of a machine learning model.
- ❖ Thus, the results can prove to be exceedingly beneficial for the agriculture farmers. We have used a dataset from Kaggle containing 2200 values of 22 unique crops.
- ❖ We have applied machine learning algorithms on the dataset. The dataset is used to train the model according to the actual values and then test the model for its accuracy.
- ❖ According to Prof. Nischitha K., presently our farmers are not using technology and analysis, so there may be a chance of wrong selection of crops for cultivation which will reduce their income.
- ❖ So, we have also included graphical user interface to make it seamless and attractive to operate for even the first-time users who have no prior experience of using any such app or facility.

- ❖ Our research aims to take into consideration the ground reality reflecting the actual requirements instead of assuming any of the factors. Since we are mainly catering to uneducated farmers, we have made it more and more visual and easier to operate.
- ❖ This process is laying the foundation for further evaluation through addition of secondary factors having an impact on deciding the crop to be grown on a particular field in a particular place.



2.1 Research Approach :-

- We have used scikit-learn to implement machine learning algorithms on the dataset. The model is trained in such a way that it learns from the data given to it and applies the same trends and knowledge to give optimum results for any given input.
- The model is then implemented using GUI after testing to give a visual look and clarity in accessing the machine learning results from the front-end.



2.2 Method Approach:-

- The dataset is loaded in the ML model. Then after splitting of data into training and testing sets, it is trained and tested using three different algorithms.
- The best algorithm is chosen and then the model is finalized. This model now performs its action by taking in different factors as input and returning the optimal crop yield.



2.3 Kaggle Dataset:-

- ✓ Kaggle is a subsidiary of Google which gives users a platform to get and publish data sets. Apart from this, it also allows the users

to build models in an environment that is generally web-based and data-science oriented.

- ✓ It also hosts competitions where people can compete and hone their ML skills and also get some useful research ideas. We have taken our dataset from Kaggle which we have used to train our model.

```
# displaying classification report
cr_NB = classification_report(Y_test, Y_prediction_RFC)
print(cr_NB)
```

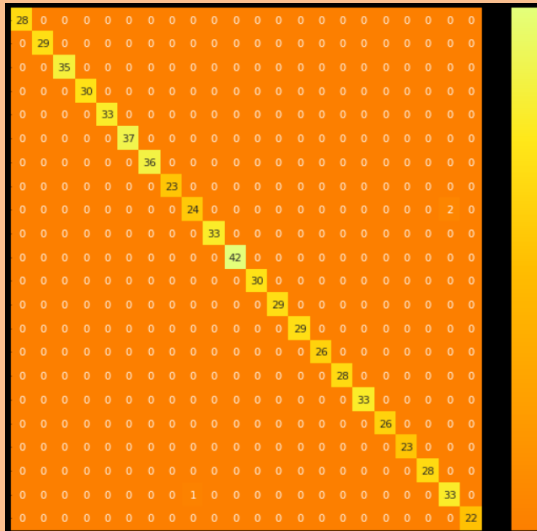
Output exceeds the [size limit](#). Open the full output data [in a file](#).

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	28
banana	1.00	1.00	1.00	29
blackgram	1.00	1.00	1.00	35
chickpea	1.00	1.00	1.00	30
coconut	1.00	1.00	1.00	33
coffee	1.00	1.00	1.00	37
cotton	1.00	1.00	1.00	36
grapes	1.00	1.00	1.00	23
jute	1.00	0.92	0.96	26
kidneybeans	1.00	1.00	1.00	33
lentil	1.00	1.00	1.00	42
maize	1.00	1.00	1.00	30
mango	1.00	1.00	1.00	29
mothbeans	1.00	1.00	1.00	29
mungbean	1.00	1.00	1.00	26
muskmelon	1.00	1.00	1.00	28
orange	1.00	1.00	1.00	33
papaya	1.00	1.00	1.00	26
pigeonpeas	1.00	1.00	1.00	23
pomegranate	1.00	1.00	1.00	28
rice	0.94	1.00	0.97	34
watermelon	1.00	1.00	1.00	22

2.4 Machine Learning:-

- We have used machine learning to make our model capable of suggesting the optimum crop which can be sown by a farmer according to various input factors.
- Machine learning is a method of analyzing data to automate the building of an analytical model. It is in fact a branch of AI as it is

based on the concept of systems learning from data and intercession.



Algorithms-

This section discusses the platform and major modules/libraries used to implement the programming of the algorithm for crop prediction through machine learning for prominent accuracy in utilization of land and crop cultivation

3.1 Scikit-Learn Scikit-learn

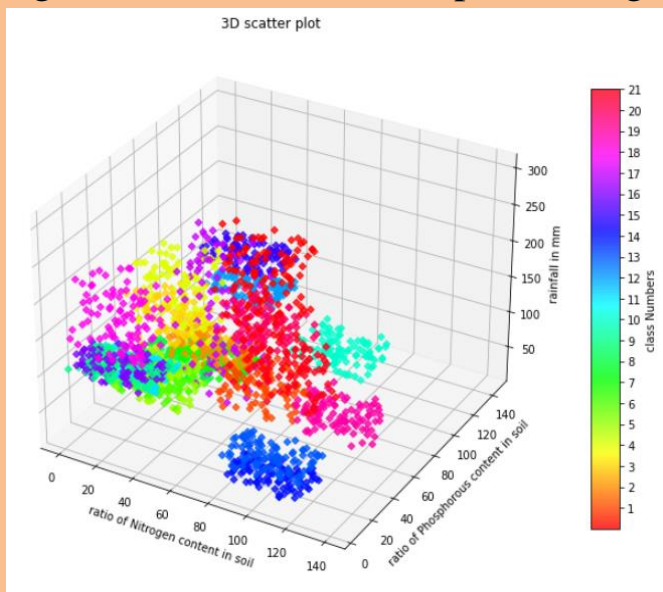
- It is an ML library for python. It has various algorithms for classification and regression like logistic regression, random forest classifiers and support vector machines.
- We can operate it along with other python libraries like NumPy and pandas. Scikit-learn is one of the best libraries especially for supervised learning which involves training the model by loading a sample dataset which it can observe and structure its learning accordingly.

- It also gives us the provision to use `train_test_split` for making training and testing datasets.

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

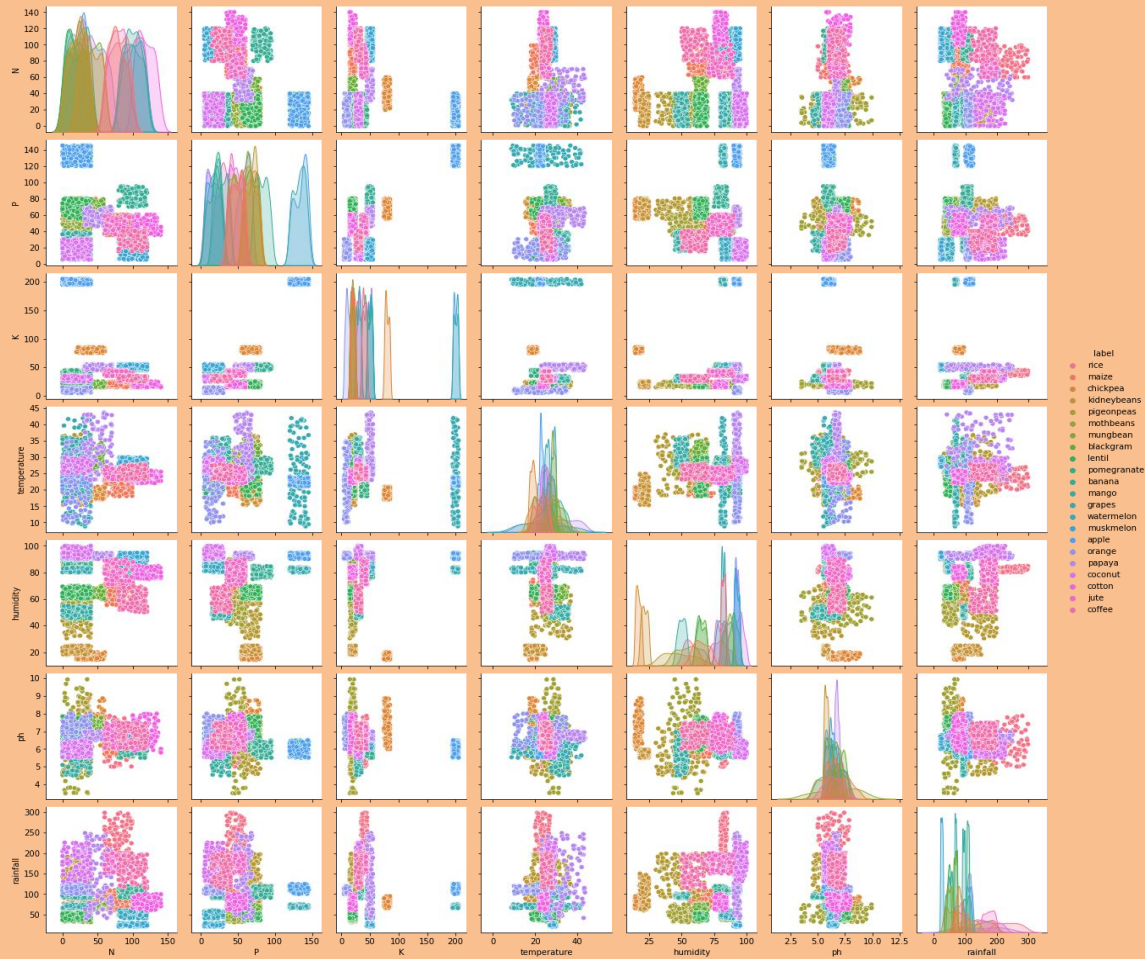
3.2 `train_test_split` Library:-

- ❖ We import `train_test_split` from `scikit-learn` to split our dataset into two different sets according to our needs- one for training the model and the other for testing the working and accuracy of the trained model so that we can choose the best possible algorithm.
- ❖ We apply different supervised learning algorithms on the training data and obtain results on testing data for all the applied algorithms. Then the best performing algorithm is selected for the



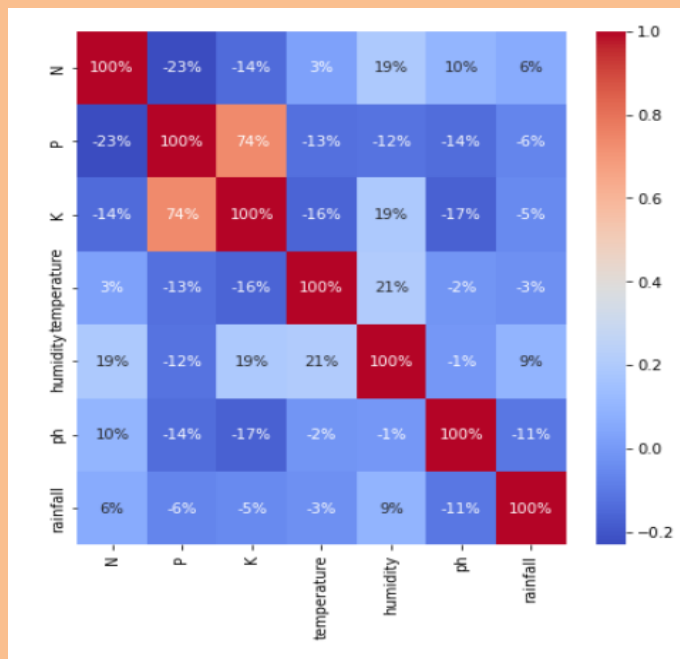
3.3 Logistic Regression:-

- Logistic regression is a basic linear model that uses a logistic function for model creation.
- It categorizes the data into discrete classes by figuring out the relationship trends from the given dataset.
- It is easy to implement and very efficient to train and can classify unknown data records considerably quickly.
- But it by default assumes a linear relation between dependent and independent variables which can turn out to be a limitation in the performance of the model in some cases.



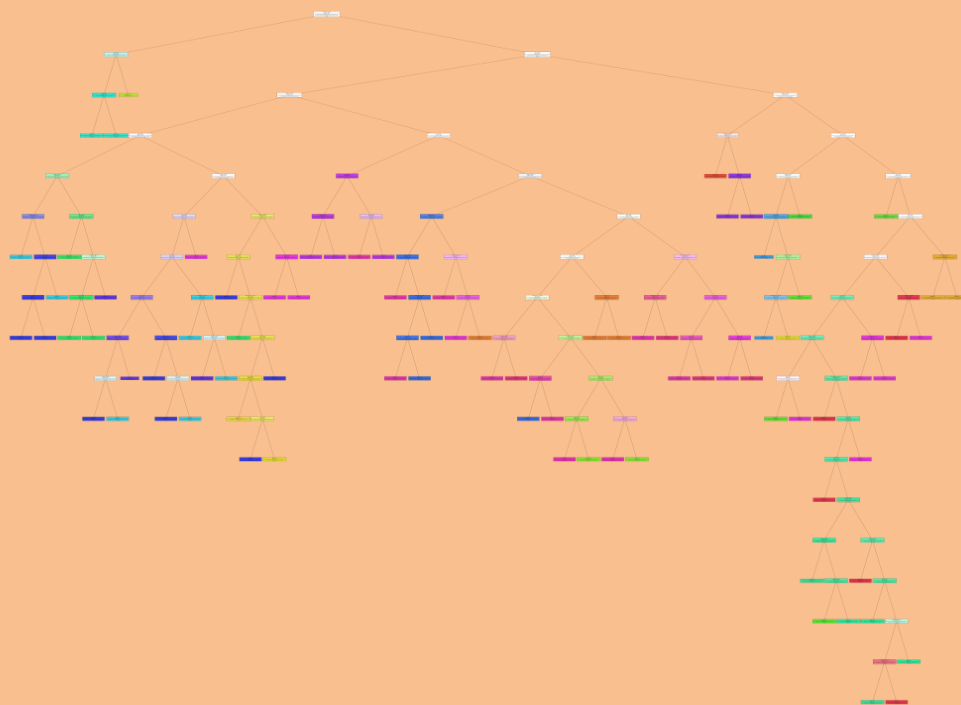
3.4 Support Vector Machines (SVM):-

- ✚ Support Vector Machines or SVM consists of a group of algorithms that analyze data for regression and classification. It represents different classes in a single plane iteratively to minimize the error and is also memory efficient.
- ✚ This makes it one of the best algorithms to use if the error persists in basic linear regression. But in case of noisy datasets and large datasets, its performance dips because of chances of overlapping of classes.



3.5 Random Forest Classifier :-

- ❖ Random forest algorithm is one of the most famous and a widely used supervised learning technique.
- ❖ It contains a number of decision trees for different subsets of the data instead of working on the whole data as a single subset.
- ❖ This improves the accuracy of prediction of the model by several folds as it takes the average of predictions of all the trees and decides the final output on the basis of majority votes of the predictions.
- ❖ This makes it suitable even for the large and varied datasets as it can deliver results with high accuracy in very less amount of time.



3.6 Working Algorithm Applied

- ❖ Jupyter notebook was used for creating the model. We started off with importing the required libraries that were pandas, matplotlib, seaborn and sklearn, which we felt were essential for carrying out proper analysis of the given

Results:

- We have used a sample data set from Kaggle, which consisted of various factors needed for the proper growing of a crop.

- We have taken into consideration these factors which will be taken as an input from the farmer/authority that will be deciding which crop to sow.
- A detailed analysis has been carried out on these factors and various inferences were generated from the results that we then took into consideration while predicting the optimum crop.

```
Ridge -> Accuracy score: 75.00 %  
RBF SVM -> Accuracy score: 98.03 %  
Nearest Neighbors -> Accuracy score: 98.33 %  
Naive Bays -> Accuracy score: 99.55 %  
Decision Tree -> Accuracy score: 98.33 %  
Random Forest -> Accuracy score: 99.85 %  
Ada Boost -> Accuracy score: 15.61 %  
QuadraticDiscriminantAnalysis -> Accuracy score: 99.09 %
```

Conclusion:-

- We have worked on a sample dataset from Kaggle which has taken into consideration records obtained from a broad agricultural demography.
- Farmers generally use hit and trial method which leads to wastage of land and resources or even disproportionate growth of crops.
- We are trying to break all such taxing walls by providing them with an accurate and justified model made by machine learning using random forest classifier to identify the correct crop to be grown in their farms.
- This will help them in improving their crop production both qualitatively and quantitatively. This will also help them to maintain the quality and nutrition contents of the soil

Future Scope:-

- As concerning future score, when the farmers sow a particular crop, there might face some issues or disease crop before harvesting. In that case, they can upload the photographs of the crop and the soil report.
- Then the AI model can identify the problems and provide them with probable solutions.
- We can also provide IOT solutions through APIs or virtual agents which can help the farmers connect with raw material dealers, who can provide them with the materials required for instance seeds and fertilizers according to the crop recommended to them by the model.

References:

[1] <https://power.larc.nasa.gov/data-access-viewer/>

[2] Anshal Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Israni, Alisha Mandholia , Sanya Bhardwaj (2015), 'Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture', Innovations in Information, Embedded and Communication systems (ICIIECS)

[3] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh (2015), 'Crop Sel Yield Rate using Machine Learning Technique', International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)