Recommendation of Paddy Seed Prediction Using Generative Model

S. Sivagurunathan[1], Thangavel Chandrakumar[2], Dhinakaran Sakthipriya[3], S. Siva Sri Varshini[4]

[1]*Gandhigram Rural Institute, India*
[2,3,4]*Thiagarajar College of Engineering, India*

Abstract: India's agriculture sector serves a significant role. It is essential to update ideas and methods in order to maintain standards in order to keep up with global warming and other demographic and materialistic issues that keep piling up. Being the top producer of a range of crops throughout the world has made it necessary to update ideas and methods in recent years. Predictive systems are a crucial instrument for management and decision-making in every productive industry. In agriculture, it is particularly valuable to have advance knowledge of a farm's profit potential. Consequently, depending on the time of year when this information is available, crucial decisions can be made that impact the farm's financial stability. The objective of this study is to develop a model for predicting Seed and pertinent characteristics in advance that is easily accessible and usable by farmers through a web application. This includes more sophisticated methods that have been derived from machine learning technologies and algorithms such as generative models and random classifiers. The fundamental motivation behind this work is to facilitate prediction using a machine learning model. The Seed suggestion model provides assistance in resolving a variety of innovative and frequently disregarded challenges that the younger generation of farmers is currently facing. Research on yield estimation in agriculture is beneficial for farmers as it helps them minimize crop loss and maximize profits by obtaining the best prices for their increased yield production. Prioritizing crop-sowing techniques in agricultural research also improves the well-being of the land, farmers, and others.

Keywords: Paddy Seed, Machine Learning, Naive Bayes, Random Forest, Decision tree, Support vector machine, XGBoost, Generative model, Streamlit.

1. INTRODUCTION

Throughout the beginning of time, agriculture has been revered as India's central cultural practice. This philosophy has its origins in India's ancient Vedic literature. Because ancient humans grew their food on their own land, the plants have adapted over time to meet the demands of their diet. As a result of technological advancements that have increased efficiency and output, modern agriculture is gradually dying out. As a direct result of these successful innovations, people are spending the vast majority of their time and effort on the development of artificial items, which are hybrid products, and this in turn contributes to an unhealthy way of life. Modern humans often aren't well-versed in the ins and outs of when and where to sow and harvest crops. Alterations in seasonal climatic conditions, brought on by these farming practices, have a deleterious effect on vital resources like soil, water, and air and raise the prospect of food insecurity.

The agriculture sector has benefited from the application of machine learning techniques, which have increased both crop yield and product quality. These techniques involve collecting specific data and using tailored algorithms to analyze factors such as soil composition, climate, terrain, and the presence of illnesses or insect infestations. By understanding how crops perform under different conditions, farmers can make more informed decisions about what to plant and how to care for their soil. Predicting nutritional deficits in advance can help maintain high agricultural output and protect the natural world.

This work is to improve the better analysis of cultivation time in order to select the better paddy seed and identify the relevant features for producing more output of crops, which should help overcome the economic condition of farmers. The purpose of this work is to improve the better analysis of cultivation time. The format of this research project is as described in the following: Section 2 contains a representation of the literature review, Section 3 contains a discussion of the materials and technique, Section 4 includes a presentation of the outcomes and discussion, and Section 5 consists of a conclusion and potential future applications.

2. RELATED WORKS

In this section, we will talk about previous study that is relevant to two things. The first one is about using agricultural data analysis, while the second one is about web-based agricultural data analysis incorporate the prediction of ML techniques.

In recent years, numerous academicians have increased their use of advanced mathematical algorithmic techniques, such as machine learning, in crop ET modeling. The constructed a model to predict plant transpiration based on plant and environmental characteristics and random forest regression [1]. The SIMDualKc model devised by While neural networks can process enormous amounts of data, some training methods, such as support vector regression, cannot manage the required volume of data. AdaBoost's sluggish training periods are due to its relatively subpar classifiers. The bagging and random forest algorithms may be susceptible to underfitting [2].

To proposed a technique for evaluating soil fertility that makes use of multiple categorization algorithms applied to the soil data set. The results of the predictions made using this strategy were very reliable. Programs like J48, the Naive Bayes approach, and the Random forest algorithm fall into this category. The J48 algorithm provided the most accurate results (98.17%) out of all the algorithms we tested [3]. The investigated of similarities and differences between three machine learning algorithms: Naive Bayes, JRIP, and J48. They did this so they could make predictions about the types of soil. The data collection and analysis took place over the course of two years, and a data analysis program was used to examine the results of 110 soil samples. Based on the findings of the evaluations, the JRIP algorithm's performance should have been higher. It achieved the highest degree of accuracy (98.18%) and a kappa statistic (0.99) that was near to value (1). To developed a strategy to raise the value and profit of agricultural land by using data mining tools to assess agricultural productivity. Using this strategy, agricultural output might be estimated through the use of data mining techniques [4].

To determine the type of rice crop production from the presence of macro and micronutrients, applied a number of machine learning algorithms. The goal was to improve the reliability of their forecasts, therefore they did this. When it comes to forecasting harvest success in relation to soil nutrients, these two models also shine [5]. To be presented a data mining approach to estimating agricultural yield for the category of soil datasets under study. It was shown that this technique might be used to predict the category. In doing so, we make use of both the Naive Bayes and K-Nearest Neighbor methods. Furthermore, data mining technology is used to build the system's design, which is subsequently evaluated [6]. The author suggested the use of fuzzy-c means clustering and neural networks for wheat crop forecasting [7]. To be used the Random Forest technique to forecast agricultural output in the Indian state of Tamil Nadu. This project was completed in the same calendar year it was initiated. The R Studio program was used to carry out the tests; the dataset for the study included a wide variety of elements, such as precipitation, temperature, crop yield, and so on [8].

Crop yield is the dependent variable, and this analysis of environmental factors shows that all of these environmental variables have an effect on crop production. Cultivation Area and the Food Price Index have less of an effect on crop productivity than weather and soil parameters [9]. To be developed a data mining-based solution to assist farmers in monitoring soil quality. The system prioritizes soil quality inspections so that it can learn which crops thrive in a given environment [10]. Using a variety of machine learning approaches, suggested a system for categorizing crop yields according to the macro- and micronutrient content of the soil. High-yield, high-profit crops are identified by this technology's analysis [11]. For this purpose, adapted the decision tree method by including soil information in the classifier. By grouping soils with similar characteristics together, farmers will have a better idea of what to expect from their plantings [12]. A developed a decision support system to boost crop yields by keeping an eye on the weather and other agronomic variables. The dirt used in the tests was from the Agra region. This system required a lot of memory and computing power, despite the high quantity of datasets [13].

The different ways to use a number of machine learning methods, such as KNN, SVM, and Decision Tree, to estimate the sugarcane crop yield. For the purpose of this study, the Python platform was utilized to carry out the research. The Decision Tree was expected to have an accuracy of 99%, which was the highest conceivable, and it also had a smaller mean square error [14]. Using Naive Bayes, Decision Tree, and a hybrid technique that combines Naive Bayes and Decision Tree, to be prescribed the choose nutrients for evaluating the soil supplements. The nutrients consisted of a wide variety of elements, some of which were nitrogen, phosphorus, potassium, calcium, magnesium, sulphur, iron, and zinc. When evaluating the efficacy of various categorization algorithms, both the degree to which they are accurate and the amount of time it takes for them to complete their assignments are taken into consideration [15]. In their research, to be utilized ML techniques to make predictions about the soil's fertility. The authors conducted research to develop a model that uses Random Forest and K-Nearest Neighbor algorithms to accurately and efficiently determine whether soil is suitable for crop production based on soil parameters [16].

This research project's primary objective is to implement and evaluate a variety of Machine Learning Algorithms using Python libraries in order to generate the most accurate predictive model feasible. This model will be used to assist farmers in finding a better way to increase their crop yield, which will depend on the level of the required Seed dataset and will enable us to select the type of seed features to help the further crop given more profit in future most suitable for the subsequent prediction process to incorporate the result in web application for this proposed work.

3. MATERIALS AND METHODOLOGY

A. *Study Area & Dataset Description*

This research focuses on the agricultural statistics of the paddy crop in the Madurai district of Tamil Nadu, India. The study collected data from various sources including the Tamil Nadu State Government's Department of Statistics and Department of Agriculture, as well as seed vendors and farmers. The data covers a twenty-year period from 2002 to 2022 and includes 18 characteristics of paddy seeds such as length, average yield, days to maturity, seed type, grain weight, color, and parentage. Table 1 provides more detailed information about the dataset.

Table 1. Paddy seed and environmental factors

S.No	**Attribute Name**	**Attribute ID**	**Description**
1	Paddy Seed Name	PSN	Ruling Varieties based on duration (Short/Medium/Long/Hybrid), Collection of Paddy Name in south districts
2	Parentage	PT	Hybrid seeds
3	Period	PD	Period (duration) (Short/Medium/Long)
4	Maximum Duration	MaxD	Duration based on no. of days Max-160 days
5	Minimum Duration	MinD	Duration based on no. of days Min-94
6	Average Yield	AY	Average Yield of Paddy (Kg/ha)
7	Grain Weight	GW	Grain weight
8	Grain Type	GT	Grain Type (Long/short/medium-Bold/Slender, small) etc.
9	Habit	HT	Habit of the Crop (Semi-dwarf/Semi-dwarf, Semi-erect/Semi-dwarf, slightly open), etc.
10	Rice Color	RC	Colour of the Rice
11	Special Feature	SF	Additional features of the paddy crop
12	Starting Month	SM	Starting month of season
13	Ending Month	EM	Ending month of season
14	Rainfall Actual	RFA	Rainfall Actual ratio
15	Rainfall Normal	RFN	Rainfall Normal ratio
16	Temperature Minimum	Tmin	Temperature max 10 – 12 degrees
17	Temperature Maximum	T max	Temperature min 36 – 38 degrees
18	District	DT	District list in South Tamil Nadu

B. *Dataset Integration*

i. Data Processing

An integral part of utilizing ML methods is the preparation of data. Data preparation is the process of organizing, standardizing, and merging data before it can be analyzed. Improving the data's quality and making it more amenable to the specific ML further procedure leads to more accurate predictions. In this pre-processing to used in the paddy seed dataset but for this model, only 18 attributes have been used.

ii. **Reducing the dataset**

In this study, we are analyse a dataset with very few missing values. The main dataset was reduced by one crop because of inaccurate information in the seed dataset.

iii. Replace missing value with mean

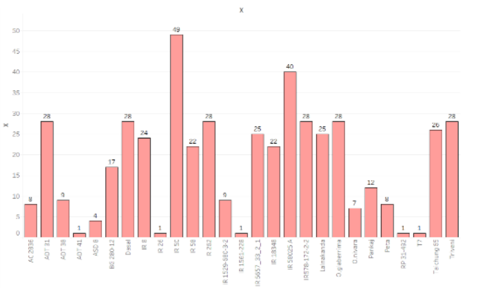
A small number of absent values were present in the dataset utilized for the paper. Therefore, the row with missing values was substituted with the average value.

iv. Encoding

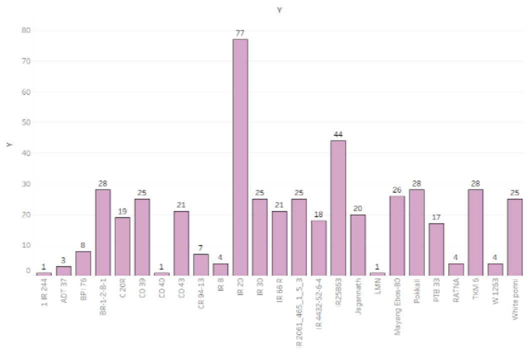
Now after we have cleaned and normalized the numerical variables present in the dataset. We have to deal with categorical variables. Nominal variables don't have any order like we have in our dataset. These two columns have to be encoded so that they can be taken as independent features in the Machine Learning Model. We have used exact encoding for both the columns. Effect/Sum encoding is like One hot encoding but instead of keeping some rows fully 0 it replaces these 0's with -1 so that we don't get a Sparse matrix. We can see after encoding both the columns we finally get 18 feature columns.

v. Exploratory data analysis for paddy seeds

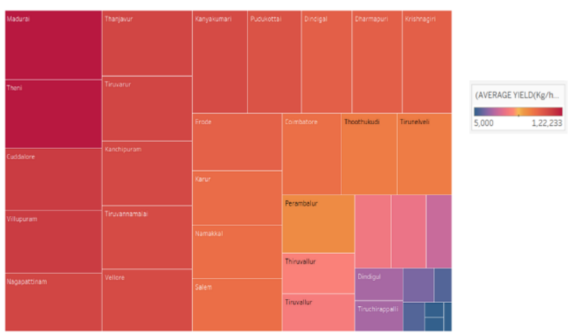
After performing cleaning and normalizing the datasets we performed Exploratory Data Analysis to get a better understanding of the agricultural scenario of India. Through plotting and analysing the graph came to know many useful information and could understand the dataset better.



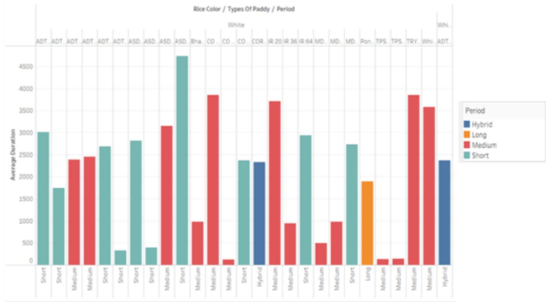
(a) Seed Value - X



(b) Seed Value - Y



(c) District based Seed production ranges



(d) Seed Time Duration ranges

Figure. 1 EDA analysis for seed factors

Fig. 1 (a), (b), (c), (d) showcases the Average yield production rate against each District name from south TamilNadu. From the given graph it is evident that various districts have the largest production levels followed by rice paddy occupying the position and paddy seed distribution levels and time duration for each seed varieties. Following the conclusion of data collection, pre-processing was carried out to convert nominal values to numeric form, impute missing data, and identify data using statistical methods. This was done in order to ensure that the data were accurate and complete. The min-max normalization that was carried out was used to perform transformations on each of the input attributes that were within the range [0, 1]. The overall approach that will be taken to complete the task that has been proposed is outlined in detail in Fig. 1.

C. Proposed Framework

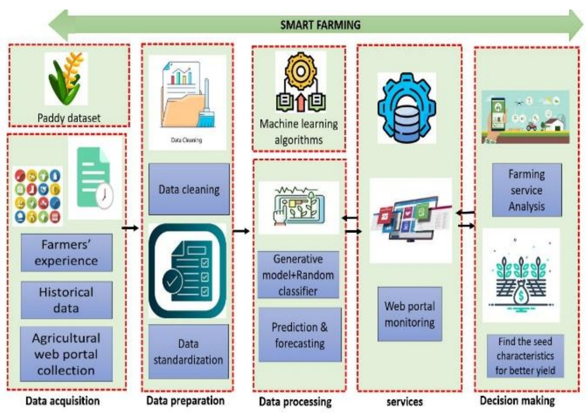


Figure. 2 Web portal Proposed Framework for Smart Farming

Fig. 2 displays the proposed system architecture and provides a copy of the suggested framework for predicting seed-specific attributes using machine learning techniques. There are essentially only three parts to the setup: (i) data gathering and first analysis ii) The machine learning process that was used to train the model, and iii) The predictions that were made. The final step is to include the forecasting outcome into the online interface for the application. This technique can be easily adapted to meet specific needs and has the potential to be used to increase crop yields on agricultural land. Information about the availability of paddy seed in each district in South Tamil Nadu must be gathered as the next phase in the data collection process. We are storing datasets obtained from various sources in our data warehouse as we acquire them. The information is then subjected to pre-processing, which includes two steps: data cleaning and data standardization. Before applying the pre-processing technique, several steps are taken with the data. When data is cleansed, it is purged of any incorrect or missing information. After that, data reduction creates a streamlined version of the digital information, typically in numerical or alphabetical form. The machine learning algorithm receives the outcomes of the data collection and data preparation steps and makes predictions based on those results. The proposed technique presented can be used to predict absolute seed features to be given growth rate in an efficient manner for increasing yield production with the help of this suggested framework, and the analyses allow this overall data to be split into train data consisting of (60%) and test data consisting of (40%) for processing with five different ML models to train and produce the various levels of accuracy us with better seed features for given good yields.

D. ML Techniques

i. Decision Tree

Decision trees are a common supervised learning technique. A dataset is segmented into smaller subsets, and decision nodes and leaf nodes are then created to form a tree-like model. The branches of decision nodes stand in for various attributes' relative importance, while the leaves stand in for potential outcomes. The root node is the most refined and advanced level of the tree.

ii. Naïve Bayes Classifier

The Bayes theorem serves as the basis for it, and it possesses autonomous properties. The estimation used for each class designation is determined by the instance's likelihood. A minimal quantity of training data is sufficient for accurately predicting the required class label for classification. When applied to data sets containing multiclass predictors, the Naive Bayes algorithm is especially useful.

iii. XGBoost Classifier

As the XG Boost Algorithm has the highest value, it is likely to be the most accurate algorithm for forecasting crop yields. The auto setting is selected by default in XGBoost so that the training process can make use of some heuristics to reduce GPU memory consumption. Because of floating point errors, their outputs could be slightly different from one another. When using this model to make a prediction, the output of each tree is first multiplied by a learning rate, and then the product of that multiplication is added to the original forecast. One is then able to arrive at a conclusive value or classification as a result of it.

iv. Support Vector Machine

Support Vector Machine is a supervised machine learning method or model that can be applied to classification and regression problems. In contrast, its primary application consists of categorization exams. The training data for an SVM are frequently represented as points in space, and these points are then divided into groups based on the maximum understandable distance between them. Each data point is an n-dimensional point in this representation, and the value of each feature is a corresponding value in this space. In order to finish the classification, it is necessary to locate the hyperplane that optimally divides the data into the two groups.

v. Random Forest

The random forest method and related random choice forest classifiers are agricultural research strategies for supervised learning that are applicable to a variety of tasks, such as division, association, and regression. Several trees are manipulated during the training and testing phases in order to determine which class most accurately represents the classification or predictive regression mode of decision trees. This process is repeated until all assignments have been produced. The random decision forest classifiers work better with individual trees than they do with their training and testing samples. Learning in this manner is referred to as supervised learning. A mapping from input to output is performed by the Random Forest algorithm. The name is inviting, and the educational principles behind it generate a number of decision trees scattered around the forest. The random forest classifiers that have a high number of decision trees are, in general, more successful than the ones that have a random appearance. In addition, the massive number of decision trees in the forest produces a staggering number of results. After that, the regression is performed with the random forest classifier. The random forest method will continue to use the numbers that are in the wrong order. In the forest, the decision tree can be found a greater percentage of the time than the unfit classifier. The advantages of the Random Forest Classifier, which were discussed previously, are also applicable to this performance indicator, for which our system makes use of the RFC algorithm.

vi) Generative Model

The application of generative artificial intelligence gives users the ability to rapidly generate new material based on a range of different inputs. Text, pictures, audio, animation, three-dimensional models, and various other kinds of data can all function as inputs and outputs for these models. A generative model is one that begins with the presumption that the signal is produced by some latent variables through the application of a transformation. A Generative Model is a potent manner of learning procedures for any form of data distribution using learning, and it has gained enormous success in just a few short years. This success can be attributed to the fact that it was developed. All different kinds of generative models have the same overarching goal, which is to learn the real data distribution of the training set so that they may produce new data points that differ in some way.

E. Working Flow

The input discusses the importance of paddy seed requirements for successful and sustainable agricultural output. It emphasizes the significance of essential seed properties and the negative impact of their absence on crop growth and yields. The input also mentions the need for studies on seed properties in Tamil Nadu and highlights the use of statistical methods and ML techniques to predict seed growth. It describes the creation of a web application using the Streamlit library in Python, which incorporates an LSTM Keras model trained with feature engineering. The app requires user input on various factors such as district, crop, soil type, season, land area, temperature, precipitation, humidity, and additional attributes. The calculated values are then used as inputs in the established model, and the website displays the predicted results. The input also mentions the installation of the local server version of the web application using the command prompt.

designed. The LSTM Keras model and weights, which are based on feature engineering, are imported and stored. This is done because our data fits this model best. In order to train the LSTM, this program requires inputs such as districts, crops, soil types, seasons, acreage, weather (including precipitation and humidity), and temperature. The built model uses those variables as inputs. The models forecast output, which is displayed on the website. The command prompt installs the web app on the local server. Functions execute web app files and launch them on the local server.

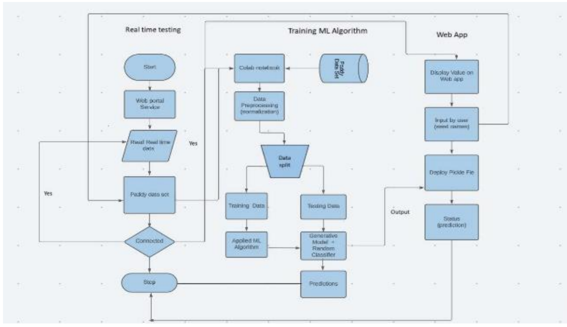


Figure. 3 Work Flow – Seed Feature Prediction

4. RESULT & DISCUSSION

A. Experiment Analysis

In this section, the results that were acquired after applying the Machine Learning algorithms to the gathered dataset are shown. These results were achieved by executing the Machine Learning algorithms. On a dataset that has been previously trained, the machine learning algorithms DT, XGB, SVM, NB, and RF have been applied. This study has made use of the Colab notebook, which is written in Python. The database of paddy seed characteristics serves as the foundation for the model that has been proposed. The above statistical report produces the most accurate value of comparison between the above ML algorithms. Fig 4. which indicates whether or not the seed features will be suitable for further yield generation. This report is generated based on the combined paddy seeds that was checked out. As a consequence of the experiments, we can observe that RF has achieved the highest possible level of precision. The accuracy of classification is summarized in the table below:

Table 2. Comparison of Accuracy Rate in ML Algorithms

ML - ALGORITHM	ACCURACY	PRECISION	RECALL	SPECIFICITY	F-SCORE

Decision Tree	85.20%	70.72%	69.33%	79.11%	0.8490
SVM	75.91%	68.61%	72.34%	69.33%	0.7812
XGBoost	70.05%	69.43%	68.11%	72.43%	0.7109
KNN	89.01%	79.57%	77.17%	80.17%	0.7915
RANDOM FOREST	94.85%	85.40%	89.70%	90.23%	0.8971

These matrices are also displayed in the bar chart located below in Figure 3 The performance metrics are displayed here in terms of accuracy, recall, specificity, and precision, as well as the F-score. On the test dataset, the accuracies of RF, NB, SVM, LR, and DT were respectively 0.99, 0.0.93, 0.89, 0.87, and 0.94; RF demonstrates superior performance in comparison to other machine learning approaches. In a manner parallel to the accuracy measure, all of the other metrics were won by RF, demonstrating that it is the superior algorithm. Additionally, DT did an excellent job. RF, on the other hand, demonstrated greater accuracy than DT. On the other hand, LR did not do particularly well when applied to our dataset. Since this is the case, it should come as no surprise that RF and DT are both effective classifiers for seed features prediction based on the various paddy seed information.

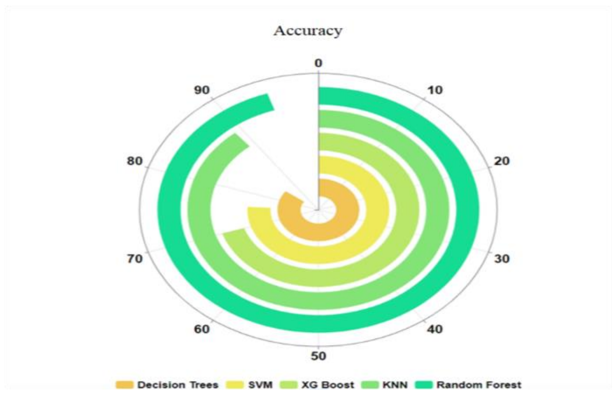


Figure. 4 Comparison of Accuracy ranges in ML algorithms

B. Training & Testing Accuracy Comparison of ML Algorithms

The accuracy of the classification that is produced by a supervised classifier is significantly impacted by the properties of the data that are utilized during training of the classifier. It is essential that the number of classes be sufficient to explain the paddy seed various features for this research area, and it is also essential that the data obtained from training and testing provide a description that is representative of each instance. In this part, the results attained by the random forest in classifying the test data given in Fig. 5 are compared with those produced by other machine learning classifiers. With the use of this comparison, we hope to find out whether the high accuracy values that we attained using the RF are dependent on the classifier. Furthermore, the research comes to the conclusion that RF classifiers are highly suggested for high-level accuracy of paddy seed data sets.

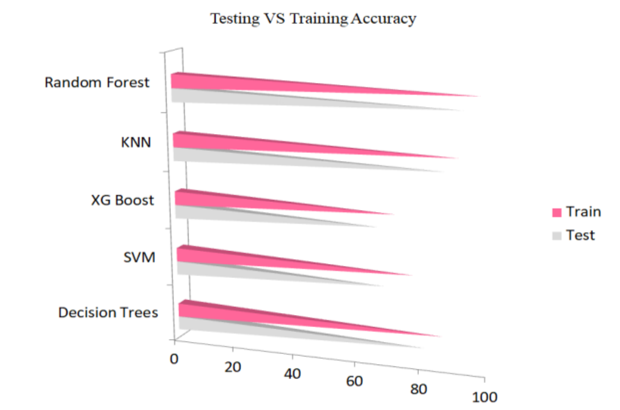


Figure. 5 Training & Testing analysis in various ML algorithms

C. Accuracy Analysis for Prediction of Seed

Due to the fact that the predicted accuracy is so high throughout both the training and testing processes, only a select handful of the samples that were used to construct the predictive models were able to be used for seed accuracy prediction perfectly. This is seen in Fig. 6, which can be found below web screens. Therefore, selecting the appropriate values for the model parameters is of the utmost significance if one wishes to avoid the learning process from improving model performance. The number of training samples was also raised, which helped improve the accuracy of the predictions. Therefore, the quantity of the dataset is a crucial component to consider while developing an improved prediction model. The following tables, which are organized according to the target variable, provide a summary of the prediction accuracies that were created during the process of model training and evaluation.

The input describes the development of a web app using the Python Streamlit module. The app utilizes a Feature engineering-based LSTM Keras model to predict crop yield and production based on various input parameters such as district, crop, soil type, season, land area, temperature, precipitation, humidity, and additional features. The model has been trained and its weights are imported and saved. The app is deployed to a local server using the command prompt, and the predicted results are displayed on the website. The deployment process is shown in Fig. 6.

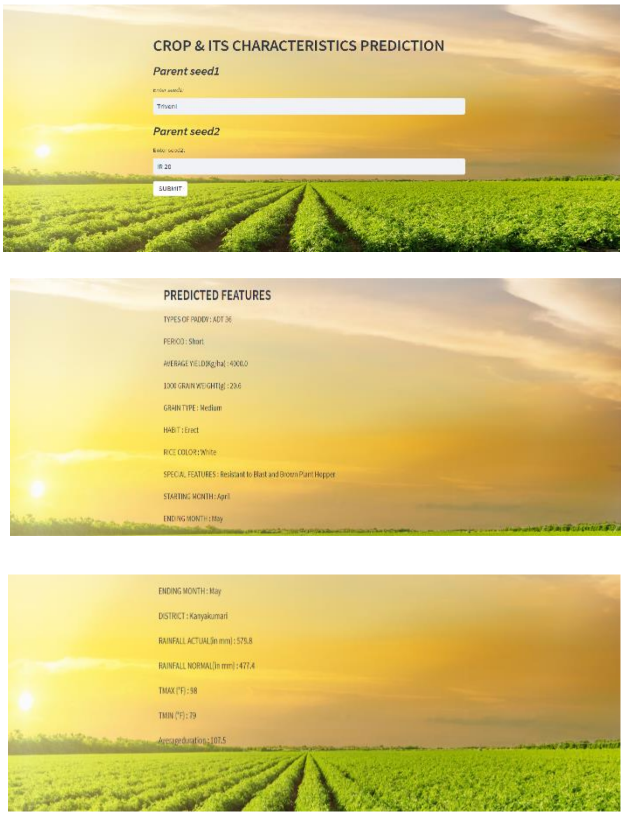


Figure. 6 ML Result incorporate Web Portal Screen

5. CONCLUSION AND FUTURE SCOPE

Additionally, the farmer can receive a projection of the crop's production in response to the crop he or she has ordered. When giving suggestions to the farmer, they take into account other crops that can be grown during that cropping season and are also good for the ratio of based on seed and environmental factors. The module for alternating crop rotation provides an additional three different plant species that are capable of being grown in that seed factors based environment. The crop suggestion method being considered offers an accuracy rate of 94.85 percent get from RF method. Following the results of the experimental study, one may come to the conclusion that ML approaches are capable of being utilized effectively for the purpose of yield prediction of seed. In contrast, the results of this research indicate that random forest is the method that provides the best accurate predictions regarding seed features. These powerful machine learning algorithms will be of assistance to farmers in the process of yield prediction in advance based on the properties of the seed combination. In the not-too-distant future, it will be possible to employ Big Data techniques for crop production prediction using massive paddy seed data sets. If there is a high crop yield projection, seed suggestions can be adopted based on the results of the yield prediction to assist seed feature analyzers and farmers in making decisions in a manner that is appropriate to the situation.

References

1.Li, L., Chen, S., Yang, C., Meng, F., & Sigrimis, N. (2020). Prediction of plant transpiration from environmental parameters and relative leaf area index using the random forest regression algorithm. *Journal of Cleaner Production*, *261*, 121136.

2.Darouich, H., Karfoul, R., Ramos, T. B., Moustafa, A., Shaheen, B., & Pereira, L. S. (2021). Crop water requirements and crop coefficients for jute mallow (Corchorus olitorius L.) using the SIMDualKc model and assessing irrigation strategies for the Syrian Akkar region. *Agricultural Water Management*, *255*, 107038.

3.Bhuyar, V. (2014). Comparative analysis of classification techniques on soil data to predict fertility rate for Aurangabad District. *Int. J. Emerg. Trends Technol. Comput. Sci*, *3*(2), 200-203.

4.Rajeswari, V., & Arunesh, K. (2016). Analysing soil data using data mining classification techniques. *Indian journal of science and Technology*, *9*(19), 1-4.

5.Singh, V., Sarwar, A., & Sharma, V. (2017). Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach. *International Journal of Advanced Research in Computer Science*, *8*(5).

6.Supriya, D. M. (2017). Analysis of soil behavior and prediction of crop yield using data mining approach. *Int J Innov Res Comput Commun Eng*.

7.Verma, A., Jatain, A., & Bajaj, S. (2018). Crop yield prediction of wheat using fuzzy c means clustering and neural network. *International Journal of Applied Engineering Research*, *13*(11), 9816-9821.

8.Priya, P., & MuthaiahM, U. (2018). Balamurugan. Predicting yield of the cropusing machine learning algorithm. *International Journal of Engineering Science Research Technology (IJESRT)*.

9.Sellam, V., & Poovammal, E. (2016). Prediction of crop yield using regression analysis. *Indian Journal of Science and Technology*, *9*(38), 1-5.

10. Ghadge, R., Kulkarni, J., More, P., Nene, S., & Priya, R. L. (2018). Prediction of crop yield using machine

learning. *Int. Res. J. Eng. Technol.(IRJET)*, *5*, 2237-2239.
11. Singh, V., Sarwar, A., & Sharma, V. (2017). Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach. *International Journal of Advanced Research in Computer Science*, *8*(5).
12. Dolas, V. M., & Joshi, P. U. (2018). A Novel Approach for Classification of Soil and Crop Prediction. *Int. J. Comput. Sci. Mob. Comput*, *7*(3), 20-24.
13. Shirsath, R., Khadke, N., More, D., Patil, P., & Patil, H. (2017, June). Agriculture decision support system using data mining. In *2017 International Conference on Intelligent Computing and Control (I2C2)* (pp. 1-5). IEEE.
14. Renuka, S. T. (2019). Evaluation of machine learning algorithms for crop yield prediction. *Int J Eng Adv Technol*, *8*(6), 4082-4086.
15. Manjula, E., & Djodiltachoumy, S. (2017). Data mining technique to analyze soil nutrients based on hybrid classification. *International Journal of Advanced Research in Computer Science*, *8*(8).
16. Jayalakshmi, R., & Devi, M. S. (2019). Relevance of machine learning algorithms on soil fertility prediction using R. *International Journal of Computational Intelligence and Informatics*, *8*(4), 193-199.

Dr.S.Sivagurunathan is an Assistant Professor in the Department of Computer Science & Applications at Gandigram Rural Institute, Dindigul, India. His research interests include Computer Networks, Network Security, Cloud Computing and IOT. He has guided many postgraduate scholars and teaches several courses on Computer science & Applications.

Dr. T. Chandrakumar is an Associate Professor in the Department of Applied Mathematics and d Computational Science (Data Science) at Thiagarajar College of Engineering, Madurai, India. His research interests include Software Engineering, Enterprise Resource Planning (ERP), and Machine Learning. He has published research papers in refereed conferences and journals such as Computer Standards & Interfaces (Elsevier-SCI), International Journal of Project Management (Elsevier-SCI), IJEIS of IGI-USA, IJICT. He has authored several Book chapters in the refereed edited book of SPRINGER.

Mrs.D.Sakthipriya is a Full-time Research Scholar in the Department of Applied Mathematics and Computational Science (Data Science) at Thiagarajar College of Engineering, Madurai, India. She is Studying PhD at Anna University, Chennai in July 2021. she holds an M.phil., from Madurai Kamaraj University, Postgraduate MCA., from Thiagarajar College of Engineering, Madurai, and Under graduate B.Sc. in Computer science from Madurai Kamaraj University, Madurai. Her research interests include Machine learning and Deep Learning. She has published research papers in refereed conferences and journals.

Ms.S.Siva Sri Varshini is a postgraduate in the Department of Applied Mathematics and Computational Science (Data Science) at Thiagarajar College of Engineering, Madurai, India. Her research interests include Machine learning and Deep Learning.