Suitable Fertilizer Recommendation System Using Linear Forest Classifier

[1]Dr. Kalli Srinivasa Nageswara Prasad, [2]S.V.V.D. Venu Gopal, [3]K. Viswa Prasad

Sasi Institute of Technology & Engineering, Andhra Pradesh, India

[1]ksnprasad@sasi.ac.in, [2]venugopal@sasi.ac.in, [3]viswa@sasi.ac.in

Abstract

Fertilizers are plant nutrients that replenish nutrients in the soil that have been depleted by crops. The fertilizers are used to meet the nutritional needs of the crop as well as to increase the nutrient content of the soil. Every crop needs a precise mix of nutrients to survive and grow properly. A farmer may have difficulties recognizing the fertilizer they require because there are so many on the market. As a result, in this instance, a fertilizer recommendation system is used. The project's recommendations are made using a hybrid method dubbed the linear Forest Classifier, which is a blend of Linear Regression and Random Forest Classifier (95.8% accuracy). The project's dataset was taken from Kaggle. Temperature, Humidity, Moisture, Soil type, Crop type, Nitrogen, Potassium, Phosphorous, and the output parameter Fertilizers are all included in the dataset. The current Recommendation Systems have various limitations in terms of the parameters included in the dataset, as well as low accuracy Scores. As a result, the project developed a recommendation system that included a suitable number of parameters and used hybrid modelling to achieve the best results.

Keywords: *Machine Learning, Recommender Systems, Linear Forest Classifier, Fertilizer Recommendation, Nutrient Classification*

1. Introduction

The great majority of Indians work in agriculture, and the country's economy relies significantly on it. As long as individuals were farming their own land, they had the freedom to experiment with a wide variety of crops. Just like people, crops throughout the globe need certain essential nutrients and minerals in order to flourish and develop strong. Food production and agricultural productivity are directly connected to fertiliser production. Fertilizers are often overlooked as a factor in food production. Soil nutrients are replenished by fertiliser use. It would be disastrous if fertilisers were not applied to crops and farms. The nutrients in the soil have a direct impact on the quality of the products that are cultivated there, which is why soil is so essential in agriculture. Nutrient diversity is critical to the health and development of any crop. Fertilizer is an essential part of a plant's growth, since much fertiliser may harm the plant and inadequate nutrient levels can lead to plant disease. There are several nutrients that may be found in the soil that can improve soil fertility and plant productivity for agricultural systems to grow and flourish [19]. Agricultural yield models, crop-specific fertiliser management, and root-cause analysis of crop stress induced by nutrient deficits all rely on soil nutrients as an important input.

Furthermore, fertiliser affects the productivity of field crops, producer earnings, and environmental quality. Since they may help growers make better agricultural choices, Fertilizer Recommendation Systems have been more popular recently. Systems that recommend products to users using a variety of factors are known as recommendation systems. Rather of sifting through a massive amount of data, the recommender system uses the user's preferences and interests as well as other information to choose just the most relevant facts. It identifies whether the user and the item are a good fit and then makes a recommendation based on the similarities between the two.

Crop-based Recommendation Systems and Fertilizer Recommendation Systems are the two types of Recommendation Engines utilised in agriculture. Crop-based recommendation systems frequently include numerous soil parameter characteristics while processing data and providing crop predictions, assuring that the crop that best matches the soil, as well as the season in which it is produced and the potential yield, is recommended. [16].

The fertilizer recommendation has various good impacts on farming; as it is known, there are numerous fertilizers on the market today, so determining which fertilizer is best for the crop is always a challenge. However, after the fertilizer has been chosen, several problems emerge regarding the amount of fertilizer that should be used to obtain the optimum outcomes, since excessive usage can be hazardous, and inadequate use might result in insufficiency [10]. Because recommending the best fertilizer encourages healthy crop growth, it leads to increased agricultural yield and makes a farmer's job easier.

In order to ensure that the crop's nutrient needs are met while also raising the nutrient level in the soil to a critical soil test level at a certain time, the fertiliser recommendations are established. Although many fertilisers are available, choosing the appropriate one for a crop is always a difficulty due to their vast variety of positive effects in farming. It's difficult to recommend a crop or season-appropriate solution using existing technologies, which have a variety of restrictions, such as the fact that they only function with certain datasets.

An answer has been found in this publication. Linear Forest Classifier, a mixture of Linear Regression and Random Forest Classifier, has been created that outperforms the current machine learning algorithms. Listed below are the sections and subsections of this document: There are similar studies in Section 2 that look at the phenomenon of spam tweets. Here, the recommended technique is laid forth. The results of our studies are summarised in Section 4. Section 5 sums up the findings and provides recommendations for future improvement.

2. Related Work

Hao Li et al. [1] introduced a series of comprehensive modeling techniques for soil nutrients using several Machine Learning Algorithms. To produce a nearly unbiased categorization estimate, cross-validation was utilized to choose smoothing parameters in the model. The approach also employed three conditions: no normalization, normalization, and normalization with normalization. Each model was expected to give precise predictions using a different normalization criterion. The disadvantage of the model is whenever employed with a big dataset, the SVM approach may not always fit, therefore target classes could overlap.

There is a system detailed in Crop Recommendation and Fertilizer Purchase System by Mansi Shinde et al. [2]. It's possible to get the recommended fertilisers just from the website itself. The user is presented with a pair of fertiliser recommendations. Soil quality is measured using the ideas of nitrogen, potassium and phosphorus levels in the soil as a framework for the study's data The dataset was used as a training set for the model before the algorithms were applied and the results were examined using historical data as a test set. Uses Random forest approach, which is 40% efficient on dataset, and ID3 algorithm (60%) to make model predictions.

An article by Aparna et al. [3] explains how supervised machine learning methods like regression, support vector machines, multi-layer perceptron, and so on might help predict agricultural fertiliser use. When it comes to pest control products, quality and proper application are key factors to consider. The WEKA tool was used to perform 10-fold cross-validation on agricultural data from a variety of crops and many machine learning methods.

According to Krutika Hampannavar et al. [4], in their work Prediction of Crop Fertilizer Consumption, they discovered that chilly plants were lacking in nitrogen, located the deficiency zone's elevation, and estimated how much fertiliser will be required by these plants. An alternative to chemical fertilisers, "biofertilizer" may give nutrients via nitrogen fixation, dissolve phosphorus, and promote plant development by synthesising growth-promoting essences, according to Kanthesh Basalingappa et al [5]. Crop production possibilities were considered, as were several formulations of these always available and environmentally acceptable nutrients.

Using a soil-based fertiliser recommendation method developed by Komal Bodake et al. [6], which can be used for regional soil analysis, farmers may better grow and produce the right crop. Farmers will be able to easily understand the tool since it will be written in their own tongue. Crop and soil conditions are recorded in a database by sensors, and this information is then used to calculate the total amount of fertiliser needed.

There are several similarities between the systems presented by Kunal Sonawanw and his colleagues and the one proposed by Kunal Sonawanw et al [7]. Extreme learning machine

settings were modified by Suchithra and Maya Pai [8] in their work to increase soil nutrient classification prediction accuracy.

By using atmospheric and soil data, Bhanumathi et al. [9] developed a system for accurately forecasting crops' output and making fertiliser ratio recommendations to end users, their article Crop Production Prediction and Efficient Use of Fertilizers contributes to increased crop yield and revenue for farmers. Classes are predicted using the Random Forest Regression method. The Random Forest and Backpropagation algorithms are used to analyse agricultural production data in a focused effort.

Machine Learning Algorithms were used by Devdatta A. Bondre and Mr. Santosh Mahagaonkar [10] to develop a system for predicting agricultural productivity based on historical data. The optimal fertiliser for each crop is recommended using machine learning algorithms like the Support Vector Machine and the Random Forest. The study's main goal was to create a model for predicting agricultural output in the future.

Anju Pratap et al. [11] attempted to examine the soil nutrient richness and assess the fertility of a particular soil sample in real time in the study Soil Fertility Analysis and Fertilizer Recommendation System. Based on the results, the system will recommend which fertilisers to use and how much to use in the soil sample to maximise production.

Soil fertility may be used to predict which crops would grow best and how much fertiliser should be administered to those crops, as shown by C.P. Wickramasinghe and colleagues [12] in their study, Smart Crop and Fertilizer Prediction System. IoT (Internet of Things) technology was utilised in a work by Lavanya et al. [13] to create a novel NPK sensor that uses LDRs and LEDs to monitor nitrogen, phosphorus, and glucose levels in soil (LED).

In the work Predicting Plant Water and Soil Nutrient Requirements, Ana Vesic et al. [14] developed an approach that would be employed in a future online platform called eAgro. It has been shown that accuracy may approach 90% when using the KNN classifier with just two features, such as the maximum and minimum daily temperatures.

According to the publication Soil Nutrients Prediction in Western India: An Evaluation of Machine Learning Models by Gunkirat Kaur et al. [15], they've devised a method for creating high-resolution soil nutrients maps while simultaneously reducing soil sampling effort and expense. Many machine learning methods were employed to evaluate data from a wide range of sources including MLR, RFR and Gradient Boosting, as well as Random Forest Regression and Support Vector Machine for Regression (SVR) (GB).

Paper Crop Yield Prediction and Fertilizer Recommendation by Varshini Naresh et al. [16] recommended the proper fertiliser ratio to increase crop yield and farmer revenue. Naive Bayes algorithm for predicting crop yields. Fertilizer Recommendation uses the KNN Algorithm, which has a high degree of accuracy. Predicting fertiliser treatment of corn using a decision tree algorithm was the subject of a study by Nusrat Jahan and Rezvi Shahariar [17].

Kanaga Suba Raja Subramanian [18] presented a persuasive approach for estimating nutrient dimension in soil and recommending adequate fertiliser in the research. The individual npk ratio will be determined subsequently using the ppm value. Prakash et al. [19] attempted to provide a strategy for predicting crop performance in a given location using data on the levels and ratios of various soil components. The system employed spatial spectrum classification, AdaBoost classifiers, and KNN algorithms, as well as Bootstrap to develop a graphical user interface, and the results were moderate.

A software based on fuzzy logic was developed by Jenskie Jerlin et al. [20] in order to provide the soil with an appropriate quantity of fertiliser. Complete, 46-0-0, Solophos, and 0-0-60 Potash Muriate are all included in the study's recommendations. Based on the specified data, a number of fertiliser ratios are created. Soil and pH samples may be used to forecast crops that are suited for the soil and fertiliser that can be used as a solution in the form of a website suggested by Preethi et al. [21] based on machine learning techniques.

Hossain and Siddique [22] presented online fertiliser recommendations for smallholder farmers in Bangladesh in order to ensure correct fertiliser consumption and enable sustainable agricultural output. The approach also demonstrated how, over time, fertiliser use grew in tandem with overall yield.

For predicting fertility and agricultural production using machine learning algorithms, the researchers Pranay Malik et al. [23] published Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield. The KNN, the Nave Bayes method, and the Decision Trees classifier were used to predict agricultural yields, and the Gini index was used to classify. Leaf Disease Prediction and Fertilizer Recommendation Engine using Machine Learning was published by Swapnil Jori et al. [24]. The use of data augmentation methods is contemplated in this case. When recommending fertilisers, the system focuses on what customers have to say rather than taking into account how different each land's nutritional needs are.

Sri Laasya.Kanuru et al. [25] presented an intelligent farming strategy that would employ a GPS module and IoT technologies to identify the nature of the soil, as well as the kind and rate of pesticides and fertilisers to be applied in an effective way. Whether k-NN is used for characterisation or relapse impacts the yield.

Hongjian Zhang *et al*. [26] stated that accurate fertiliser strength prediction can reduce the crushing rate during transportation and utilisation, ensuring efficient fertiliser utilisation, and achieving sustainable and clean crop production. For the first time, the support vector machine and the enhanced differential evolution method are used to suggest a fertiliser strength prediction model based on form attributes.

Using data from the cloud, Pandi Selvi and Poornima [27] proposed a system that could analyse the soil type, disease in the leaves, and finally recommend the appropriate fertiliser to farmers. With Palaniraj et al. [28]'s framework for crop yield forecasting and suggestions for the end-fertilizer user's ratio based on the land's atmospheric and soil specifications, crop yields can be increased and farmer revenue increased. The SVM Classifier employed in the research has a classification accuracy of more than 90%.1 percent.

Supriya and Nagarathna [29] proposed to construct a machine learning model for crop and fertilizer recommendations system based on soil features. Crop suggestions are made using the LVQ and Nave Baye’s algorithms, while fertilizer recommendations are made using the KNN classifier. The system presents the findings of a study on machine learning methodologies and compares them to neural networks to predict the best crop suggestions.

A system based on evolutionary computation was described by Usman Ahmed et al. [30] and uses an improved genetic algorithm (IGA) to make nutrient recommendations based on time-series sensor data and suggest different crop settings. This is followed by an explanation of an approach based on local knowledge for managing exploration and exploitation in order to maximise yield.

3. Architecture

A Hybrid Model, a mix of linear regression and random forest classifier, is used in this research to recommend fertilisers utilising characteristics such as temperature, humidity, moisture and soil type as well as the output parameter fertiliser name (nitrogen, potassium and phosphorus)... Libraries such as scikit-learn and pandas are used to simulate the suggested task. The main objective is to suggest the optimal fertiliser for the crop and the soil.

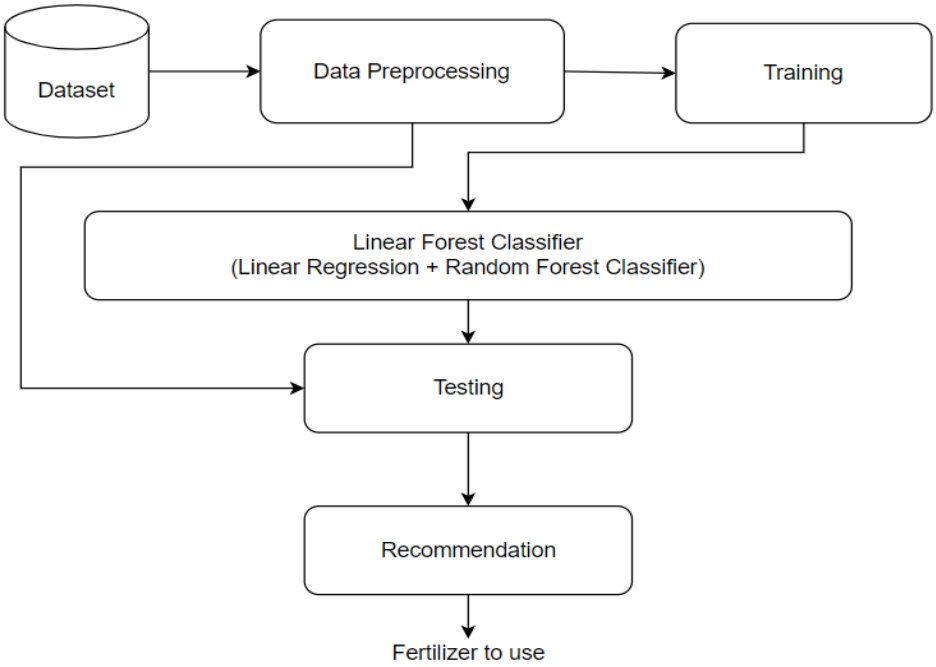

Fig. 3 Flow Diagram for Fertilizer Recommendation System.

4. Methodology

The suggested work is divided into four major modules: dataset collection, dataset pre-processing, dataset splitting, and application of machine learning techniques. The dataset is first obtained from Kaggle.

- Dataset collection
- Pre-processing of Dataset
- Splitting the dataset
- Fertilizer Recommendation model

Data Collection

What we call "data" is anything that has been structured in a certain manner. Obtaining, measuring, and analysing accurate data from several pertinent sources is the process of gathering data in order to solve research issues, answer questions, assess results, and anticipate trends and probability. In the data collecting process, the selection of high-quality data is necessary. Taken from Kaggle, this dataset includes characteristics such as Temperature (detailed), Humidity (detailed), Water (detailed), Soil Type (detailed), Plant Type (detailed), and Nutrient Composition (detailed). Data for training and testing the model includes 1089 tuples.

Data pre-processing

Data pre-processing is the process of transforming unprocessed data into something usable. Make sure the data you're utilising for machine learning or data mining is of high quality. It's possible to remove unneeded data from a dataset by data cleaning and pre-processing, resulting in a more useful dataset for data mining that contains just the most critical information. A dataset with no missing or null values was used in this study.

Label Encoding

In order for machines to read labels, the procedure known as "label encoding" must be used. Using machine learning methods, these labels may subsequently be employed in a more efficient manner. When it comes to structured datasets, pre-processing is critical. We have features like soil type, crop type, and fertiliser type in the dataset that need to be transformed to a numerical format for further analysis and interpretation.

```
In [72]: from sklearn.preprocessing import LabelEncoder
         encode_soil = LabelEncoder()
         data.Soil_Type = encode_soil.fit_transform(data.Soil_Type)
         encode_crop = LabelEncoder()
         data.Crop_Type = encode_crop.fit_transform(data.Crop_Type)
         encode_ferti = LabelEncoder()
         data.Fertilizer = encode_ferti.fit_transform(data.Fertilizer)
```

Fig. 2 Label Encoding

Splitting of Dataset

The train-test split is a method for evaluating the performance of a machine learning system. It may be used for classification or regression problems and for any supervised learning approach. The approach entails dividing a dataset into two halves. In order to build a model, the initial subset of data is utilised for training. Instead of training the model with the second subset of data, the model is given the dataset's input element and produces predictions and compares those predictions to the predicted values. To train the model, the second subset is not utilised; instead, the model is fed the dataset's input element, and predictions are created and compared with predicted values. 'Test dataset' denotes the second collection.

Fertilizer Recommendation Model

Linear and Random forest regression are utilised to prescribe fertilisers in the proposed model Using many models, such as classifiers or experts, to solve a computer intelligence issue is called ensemble learning. A frequent use of ensemble learning is to enhance a model's performance (classifications, predictions, approximations of functions, etc.) or to lessen the likelihood of an unintentional bad model selection.

Ensemble learning may also be utilised to establish a level of confidence in the model's choice. Random Forest is a well-known method that has been shown to perform well in both regression and classification applications.

Random Forest's success may be disguising its limits. It is well accepted that taking a sample from an under-explored area during training might result in low-quality predictions. In

many real-world applications, data that is not steady and falls outside of the fitting ranges is a common occurrence. Because of their potential to acquire trending characteristics, linear models may be just as successful as tree-based models in understanding a non-stationary system. They may be used to build a model capable of learning both linear and complicated connections.

Random Forest Classifier

Bootstrap and Aggregation, also known as bagging, is a technique that can be used to combine several decision trees into a Random Forest, which can then be used to solve regression and classification problems. Rather than relying just on a single decision tree, the core premise is to employ several trees to decide the final result. The ensemble learning strategy is used to classify and predict using Random Forest, a Supervised Learning technique. Many decision trees are built throughout the learning process and the output is the class that is representative of a group as a whole, such as a classification, or a mean prediction (such as regression).

There are various helpful adjustments that may be made to a random forest meta-estimator (i.e., one that combines the results of several predictions). Each node can only split a specific amount of the entire number of attributes (which is known as the hyper parameter). An ensemble model can't rely too strongly on any one feature, and all possible predictive factors are used equally.

In order to avoid over fitting, each tree generates a random sample from the original data set while creating splits. To prevent the trees from getting too entwined, the changes mentioned above are necessary. By combining the output of these decision tree classifiers, a random forest ensemble may be generated.

Consider features x1 and x2 as the horizontal and vertical axes of the decision tree outputs described above. Using model votes or averages, these results are combined into an ensemble model that outperforms the results of any single decision tree. As far as learning algorithms go, it's one of the most accurate available. For a wide range of data sets, it produces a highly accurate classifier.

Even with large datasets, it still does a good job. It doesn't have to delete any of the input variables it receives. The categorization is done by determining which variables are important. An internal, unbiased estimate of the generalisation error is generated as the forest grows. Because of this, it is very accurate even when a large percentage of the data is missing. Random Forest can be used to solve both classification and regression problems at the same time. Massive datasets with a high dimensionality can be processed with ease. Over fitting is avoided, and the model is more accurate.

Linear Regression

It is possible to determine the linear relationship between a target and one or more variables through linear regression. There are two types of linear regression, simple and multiple: Using a simple linear regression, it is possible to determine the relationship between two variables. Predictors and responses are two distinct concepts. In contrast to deterministic relationships, statistical correlations are uncovered here. The relationship between two variables is said to be deterministic if one variable can be precisely described by the other.

The key idea here is to pick the line that most closely matches the data. Prediction error (across all data points) is a measure of how well the best-fitting line fits the data (2). The equation for the line is written as eq (1). The error is defined as the difference between a given point and the regression line.

$$Y(Pred) = b0 + b1*x \qquad \text{Equation (1)}$$

It is imperative that the b0 and b1 values from eq (1) be chosen so that the error is minimised. Models are evaluated based on the sum of squared errors. The goal is to find a line that minimises these errors.

$$\text{Error} = \sum_{i=1}^{n} (\text{Actual output} - \text{predicted output}) **2 \quad x \qquad \text{Equation (2)}$$

Unless we square the error, the positive and negative points will cancel one another.
In the case of a single predictor model,

$$b0 = \bar{y} - b1\ \bar{x} \quad x \qquad \text{Equation (3)}$$

From eq (3), A positive correlation exists between the predictor and the target if b1 is greater than zero. That is to say, an increase in x will lead to an increase in y.

A negative relationship exists between x (predictor) and y (target) if b1 is less than 0. In other words, a greater increase in x will reduce y.

For example, if the model doesn't include x=0, the prediction with b0 is meaningless. When it comes to height and weight, for example, we have a dataset that indicates the correlation (y). Because height and weight can never be zero in real-time, using x=0 leads in an equation with just b0 as a value, which is completely pointless. This happened because the model's values were extended beyond their scope. It is the average of all predicted values for x=0 in the model.

In certain cases, zeroing out all of the predictor variables is challenging. This means that the discrepancy between anticipated and actual results should be as little as feasible when dealing with linear regression. The least level of error will be found in the line that fits best. To find the best fit line, we need to figure out the best weights or coefficients for the lines (a0, a1), hence the cost function is utilised to figure out the numbers.

Different regression lines may be generated by varying weights or coefficients of lines (a0, a1) by applying the cost function to the estimation of coefficient values for the best fit line.

The regression coefficients or weights are maximised by using the cost function. A linear regression model's performance is evaluated here. Using the cost function, you may check whether or not the mapping function, which converts the input variable to its output equivalent, is valid.

The Hypothesis function is a common name for this mapping function. The residual is the discrepancy between the predicted and actual figures. The residual and, by extension, the cost function will be big if the data points are far from the regression line. The residual and,

by extension, the cost function will be small if the scatter points are close to the regression line. Gradient descent is used to minimise MSE by determining the cost function's gradient.

Linear regression is based on the premise that the dependent and independent variables are connected in a straight line. It is a sign of significant correlation between the independent variables that multicollinearity is defined as. Because of multicollinearity, it may be difficult to determine the true relationship between predictors and target variables. In other words, determining which predictor variables have an effect on the target variable is tough.

It is possible to have homoscedasticity if the error term for all independent variables is the same. The error component must have a normal distribution for linear regression to work.

Confidence intervals will become either too big or too small if the error factors are not normally distributed, making it difficult to determine coefficients. The mistake is evenly distributed if the plot shows a straight line with no fluctuation. The model's accuracy will suffer substantially if there is any correlation in the error term. Autocorrelation occurs when there is a correlation between residual errors.

Many Linear Regression is a statistical method for predicting the value of a numerical dependent variable by taking into account the effects of multiple independent variables. The best possible values for a0 and a1 are found using the cost function, resulting in the best line to match the data points.

The line equation is given by eq (4) and here y is the dependent variable and x is the independent variable and m, c are the model parameters.

$$y = mx + c \; x \qquad \text{Equation (4)}$$

$$\mathbf{1/N} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (yi - (mxi + b))\ \hat{}2 \; x \qquad \text{Equation (5)}$$

In order to assess the mapping function's correctness, the cost function must be applied to the input variable. It's called the Hypothesis function, and it's a mapping function. It is necessary to utilise linear regression to determine the mean squared error (MSE) cost function, which is an average of the squared errors that occurred while comparing the predicted and actual values (5).

Linear Forest Classifier

Linear Forest classifier is an ensemble model of Random Forest Classifier and Linear regression. By integrating Linear Models with the same Random Forests, Linear Forests generalise the well-known Random Forests. The core concept is to leverage the power of Linear Models to improve the nonparametric learning capabilities of tree-based algorithms.

After fitting a Linear Model on the entire dataset, a Random Forest is trained on the same dataset using the residuals from the previous stages as the goal. The final predictions are the sum of the raw linear forecasts and the Random Forest residuals. As it is known the disadvantage of Random forest algorithm which is not being better with linearity, it often fails in the learning the cases of linear relationships among the data. so, there is one thing that can be done to make this disadvantage disappear that is to combine it with a linear model which gives it the capability to solve both linear and complex problems by eliminating its

disadvantage. Here in our proposed system this Linear Forest Classifier is used for the prediction of fertilizers.

The linear Regression merged with random Forest can only solve binary class problems so to solve multi class we have to use Linear forest classifier with OneVsRestClassifier to make multiclass Predictions. The one-vs-the-rest (OvR) multiclass technique, also known as one-vs-all, involves fitting one classifier per class.

It is compared to other classes with each classifier. Because each class is represented by a single classifier, examining that classifier may provide information about that class. This is a nice place to start when learning about multiclass classification. It is possible to use one Vs RestClassifier to categories many labels

5. Results and Discussion

This section discusses the evaluation results and experimental details of proposed model.

Linear Model

After training the dataset and applying Linear Model, we get classification accuracy of 92.2%. The Accuracy Score and Confusion Matrix are shown below.

col_0	0	1	2	3	4	5	6
Fertilizer							
0	13	0	1	0	0	0	0
1	0	38	0	0	0	0	0
2	3	0	9	0	0	0	2
3	1	0	0	26	0	0	0
4	1	1	0	1	38	4	1
5	0	0	0	0	0	29	0
6	0	1	0	0	0	1	48

Fig. 3 Linear Model Confusion Matrix

```
print(accuracy_score(y_test,logreg))
```

```
0.9220183486238532
```

Fig. 4 Accuracy score of Linear Model

Random Forest Classifier

After training the dataset and applying Random Forest Classifier, we get classification accuracy of 95.8%. The Accuracy Score and Confusion Matrix are shown below.

Fertilizer							
0	14	0	0	0	0	0	0
1	0	38	0	0	0	0	0
2	0	0	12	0	0	0	2
3	1	0	0	26	0	0	0
4	1	1	0	1	42	0	1
5	0	0	0	0	0	29	0
6	0	1	0	0	0	1	48

Fig. 5 Random Forest Classifier Confusion Matrix

```
In [48]: print(accuracy_score(y_test,pred_rand))
         0.9587155963302753
```

Fig. 6 Accuracy Score of Random Forest Classifier.

Linear Forest Classifier

After training the dataset and applying Linear Forest Classifier, we get classification accuracy of 95.8%. The Accuracy Score and Confusion Matrix are shown below.

```
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
print(classification_report(y_test,litc))
```

```
              precision    recall  f1-score   support

           0       0.88      1.00      0.93        14
           1       0.95      1.00      0.97        38
           2       1.00      0.86      0.92        14
           3       0.96      0.96      0.96        27
           4       1.00      0.91      0.95        46
           5       0.97      1.00      0.98        29
           6       0.94      0.96      0.95        50

    accuracy                           0.96       218
   macro avg       0.96      0.96      0.95       218
weighted avg       0.96      0.96      0.96       218
```

```
print(accuracy_score(y_test,litc))
```

```
0.9587155963302753
```

Fig. 7 Accuracy Score of Linear Forest Classifier

col_0	0	1	2	3	4	5	6
Fertilizer							
0	14	0	0	0	0	0	0
1	0	38	0	0	0	0	0
2	0	0	12	0	0	0	2
3	1	0	0	26	0	0	0
4	1	1	0	1	42	0	1
5	0	0	0	0	0	29	0
6	0	1	0	0	0	1	48

Fig. 8 Linear Forest Classifier Confusion Matrix.

This model Obtained the Accuracy source of 95.87%, which proves to be more than the already existing systems. The comparison with the existing models is shown below.

Model	Proposed Methodology	Accuracy
Existing Model	Naïve Bayes, Decision Tress	84.23%
Existing Model	Support Vector Machine	90%
Existing Model	NPK sensor based with Random Forest	90.60%
Existing Model	Extreme Learning Machine (ELM) with different activation functions	90%
Proposed Model	Linear Forest Classifier	95.60%

Table 1: Results Comparison

6. Conclusion and Future Enhancement

This project presented a Fertilizer Recommendation System model with the help of a Hybrid model named as Linear Forest Classifier, in which Linear Regression and Random Forest Classifiers are used in combination. This Model is able to predict the better fertilizer to use based on the Input Parameters. Initially a Linear Model is implicated on the dataset, the residuals are obtained. After that the Random Forest Classifier is trained on the residuals created on top of the dataset and the final results obtained are a combination of result from both the models. This model Obtained the Accuracy source of 95.87%, which proves to be more than the already existing systems. The Fertilizer Recommendation Model is useful for Farmers in selecting the best fertilizer from a huge number of options and helps increasing the nutrient level in both the soils as well as plant and helps for its healthier growth.

In future this project can be extended using a large data set. The dataset currently used for implementation includes 1089 tuples, which can be increased to get better results. Additional parameters related to micro and macro nutrients such as calcium, Magnesium, sulphur, Lime, carbon etc. can be included to improve the accuracy score. Several Characteristic features like leaf colour, thickness can be added to provide high end results

References

[1] Hao Li, Weijia Leng,Yibing Zhou, Fudi Chen,Zhilong Xiu, and Dazuo Yang, "Evaluation Models for Soil Nutrient Based on Support Vector Machine and Artificial Neural Networks" The Scientific World Journal Volume 2014.

[2] Mansi Shinde , Kimaya Ekbote , Sonali Ghorpade ,Sanket Pawar ,Shubhada Mone "Crop Recommendation and Fertilizer Purchase System" International Journal of Computer Science and Information Technologies, Vol. 7 2016.

[3] H D Aparna, Dr. Kavitha K S, and Dr. Kavitha, "Use of Machine Learning Techniques to Help in Predicting Fertilizer Usage in Agriculture Production" International Research Journal of Engineering and Technology (IRJET) 2017.

[4] Krutika Hampannavar, Vijay Bhajantri, and Shashikumar.G.Totad "Prediction of Crop Fertilizer Consumption" IEEE 2018.

[5] Kanthesh M Basalingappa, Raghu Nataraj, and Gopenath Thangaraj "Biofertilizer for crop production and soil fertility" Academia Journal of Agricultural Research, August 2018.

[6] Komal Bodake , Rutuja Ghate , Himanshi Doshi , Priyanka Jadhav and Balasaheb Tarle "Soil based Fertilizer Recommendation System using Internet of Things" MVP Journal of Engineering Sciences, ISSN, 2018.

[7] Mr. Kunal Sonawane, Mr. Abhishek Patil, and Mr. Sandesh Nikam "A review: intelligent fertilizers & pesticide advisory system using IOT and android" International Journal for Research in Engineering Application & Management (IJREAM), 2018.

[8] M.S. Suchithra and Maya L. Pai "Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters" Information Processing in Agriculture, 2019.

[9] S.Bhanumathi, M.Vineeth and N.Rohit "Crop Yield Prediction and Efficient use of Fertilizers" International Conference on Communication and Signal Processing, 2019.

[10] Devdatta A. Bondre and Mr. Santosh Mahagaonkar "Prediction of Crop Yield and Fertilizer Recommendation using Machine Learning Algorithms" International Journal of Engineering Applied Sciences and Technology, Vol. 4, Issue 5 2019.

[11] Anju Pratapa , Namitha Josepha , Reenu K Eapena , Rony Sebastiana , and Shane Thomasa "Soil Fertility Analysis and Fertilizer Recommendation System" International Conference on Advancements in Computing & Management (ICACM-2019).

[12] C.P.Wickramasinghe , P.L.N.Lakshitha, H.P.H.S.Hemapriya , Anuradha Jayakody, and P.G.N.S.Ranasinghe "Smart Crop and Fertilizer Prediction System" International Conference on Advancements in Computing (ICAC) 2019.

[13] Lavanya Ga, Rani Cb, and Ganeshkumar P "An automated low cost IoT based Fertilizer Intimation System for smart agriculture" ELSEVIER 2019.

[14] Ana Vesić, Vuk Ignjatović, Sava Lakićević, Luka Lakićević, Bojan Gutić, Hristo Skačev, Dušan Dotlić, Andrej Mićović, Marina Marjanović Jakovljević, and Miodrag Živković "Predicting Plant Water and Soil Nutrient Requirements" IEEE 2020.

[15] Gunkirat Kaur , Kamal Das , and Jagabondhu Hazra "Soil Nutrients Prediction using Remote Sensing Data in Westren India :An Evaluation of Machine Learning" International Geoscience and Remote Sensing Symposium, IEEE 2020.

[16] Varshini Naresh, Vatsala. B. R, and Dr. C. Vidya Raj "Crop Yield Prediction and Fertilizer Recommendation" Volume 10 Issue No.6, IJESC, 2020.

[17] Nusrat Jahan, and Rezvi Shahariar "Predicting fertilizer treatment of maize using decision tree algorithm" Indonesian Journal of Electrical Engineering and Computer Science Vol. 20, No. 3, 2020.

[18] Dr.S.UshaKiruthika, Dr.S.Kanaga Suba Raja Subramanian, S.R.Ronak,, S.Rengarajen, P.Ravindran " Design and Implementation of Fertilizer Recommendation System for Farmers" ISSN 2020.

[19] Prakash U.M, Sristika Bora, Abhishek Gautam, and K.R. Gokul Anand "Knn based Crop and Fertilizer Prediction", International Journal of Engineering and Advanced Technology (IJEAT) ISSN Volume-9 Issue-4,2020.

[20] Jenskie Jerlin I. Haban , John Carlo V. Puno , Argel A. Bandala , Robert Kerwin Billones , Elmer P. Dadios ,and Edwin Sybingco "Soil Fertilizer Recommendation System using Fuzzy Logic" IEEE 2020.

[21] Preethi G, Rathi Priya V, Sanjula S M, Lalitha S D, and Vijaya Bindhu B, "Agro based crop and fertilizer recommendation system using machine learning" European Journal of Molecular & Clinical Medicine, Volume 7, Issue 4, 2020.

[22] Hossain, M.A. and Siddique, M. N. A, "Online Fertilizer Recommendation System (OFRS): A Step Towards Precision Agriculture And Optimized Fertilizer Usage By Smallholder Farmers In Bangladesh" European Journal of Environment and Earth Sciences, 2020.

[23] Pranay Malik, Sushmita Sengupta, and Jitendra Singh Jadon, "Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield using Machine Learning Algorithms", IEEE 2021.

[24] Swapnil Jori , Rutuja Bhalshankar , Dipali Dhamale, and Sulochana Sonkamble, "Healthy Farm: Leaf Disease Estimation and Fertilizer Recommendation System using Machine Learning", International Journal of All Research Education and Scientific Methods (IJARESM), Volume 9, Issue 4,2021.

[25] Sri Laasya.Kanuru, Amit Kumar Tyagi, Aswathy S U, Terrance Frederick Fernandez, N Sreenath and Shashvi Mishra, "Prediction of Pesticides and Fertilizers using Machine Learning and Internet of Things", International Conference on Computer Communication and Informatics (ICCCI) 2021.

[26] Hongjian zhang, Chunbao xu, and jinxing wang, "Fertilizer Strength Prediction Model Based on Shape Characteristics", voulume 9, IEEE 2021.

[27] Dr.P. Pandi Selvi ,and P. Poornima, "Soil Based Fertilizer Recommendation System for Crop Disease Prediction System", International Journal of Engineering Trends and Applications (IJETA)- Volume 8 Issue 2,2021.

[28] Palaniraj A, Balamurugan A S, Durga Prasad R, and Pradeep P, "Crop and Fertilizer Recommendation System using Machine Learning", International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 04, 2021.

[29] Supriya M.S. , and Nagarathna, "A Machine Learning Based Crop and Fertilizer Recommendation System", International Journal of Computer Sciences and Engineering, ISSN Vol. 9, Issue.7, 2021.

[30] Usman Ahmed , Jerry Chun-Wei Lin , Gautam Srivastava, and Youcef Djenouri, "A nutrient recommendation system for soil fertilization based on evolutionary computation", ELSEVIER, 2021.