An Innovative Approach for Fruits Farming Recommendation System Using ML

Falguni Adhikary[1], Swarup Kumar Paul[2], Sandip Saha[3], Arindam Mondal[4], Anirban Mukherjee[5]
[1,2,3,4,5]*RCC Institute of Information Technology, Kolkata, India*

***Abstract*- Fruit cultivation contributes significantly to the agricultural GDP of India. Fruits like mangoes, bananas, grapes, citrus fruits, and apples are among the major cash crops, generating substantial revenue for farmers and the economy. Researchers are trying to developed new technique for fruit cultivation in any climate change [1]. This paper proposed a new recommendation model using machine learning algorithm[1] [ML] and voting classifier [2] for fruit farming in different climates and lands of India. The proposed model can predict the profitable fruit to grow on any farming land on the basis of soil parameters and environmental parameters. It not only reduces the risk factor of loss but also increases the profit.**

Index Term— fruits farming, machine learning, hard Voting classifier.

I. INTRODUCTION

In present agriculture scenario the fruit cultivation is one of the major factors for Indian economy. Fruit cultivation provides employment opportunities to a large number of people, both directly and indirectly, in various stages such as cultivation, harvesting, transportation, processing, and marketing. India also tries to grow fruit which include apricot, starfruit, ice apple, finger lime, pineapple etc.

The proposed model will suggest the profitable fruits to cultivate on the basis of local soil parameters and relevant environmental factor. The fruit cultivation in the Indian agricultural landscape not only contributes significantly to the economy but also serves as a vital source of nutrition, employment, and sustainability. Addressing challenges and leveraging opportunities in natural fruit farming can further enhance its positive impacts on fruit farming and overall socio-economic development in India.

Algorithms such as K-Nearest Neighbour (KNN), Logistic Regression with PCA, and Hard Voting Classifier are used to make more accurate prediction. Using this advanced techniques it increases fruit yields, improves resource management & greater profitability for farmers.

II. PROBLEM STATEMENT

The proposed Fruit farming recommendation system provides are data driven understanding of which fruit are based for local area of the farmer.

Collecting relevant and accurate information of soil parameter is vital to accurate prediction. This model will analysis the data on land condition and weather condition to recommend the right fruit farming. The problem analysis is demonstrated in Fig 1.

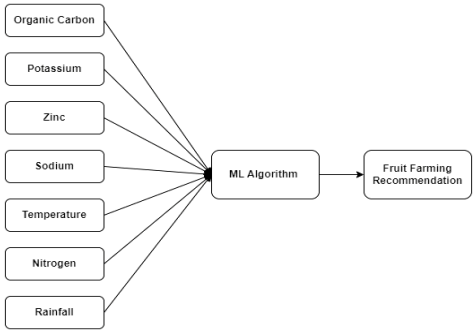

Fig 1: Problem Analysis

K-Nearest Neighbour(KNN), Logistic Regression with PCA, and Hard Voting Classifier are introduce to analysis the dataset for training and testing to recommend the fruit farming.

III. LITERATURE STUDY

In advancement of new classification model more researcher are now developed new technology/ model for fruit recommendation for farming.

P. Devika,Suwarna Gothane; Degala Divya Priya used SVM [1] to detect the diseases of fruits in early stages of orange fruit disease & model accuracy is 96.66667%.

Khalied Albarrak,Yonis Gulzar,Yasir Hamid, Abid Mehmood, Arjumand Bano Soomro suggested proposed model using deep learning technology for date fruit classification using CNN [2] with 99% accuracy based on MobileNetV2 architecture.

J. Zhao, J. Tow and J. Katupitiya presented an algorithm on vision based [3] on tree fruit recognition like apple located in single image through texture properties and color data.

L. Hou, Q. Wu, Q. Sun, H. Yang and P. Li, used to make fruit recognition based on convolution neural network[4] for fruit recognition. The selective search algorithms are used to extract the image region and it's used as input for CNN to train and recognition. The final decision is made from the basis of voting mechanism.

N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar and N. K. Cauvery used Random Forest, Naive Bayes, and Linear SVM as ensemble model [5] to make crop recommendation. Major voting is used to combine on individual basis. The model classification accuracy is around 99.91% .

IV. PROPOSED SYSTEM

The proposed system consisting of five stages that are demonstrated in work flow diagram (Fig 2) & that are dataset preparation (fruit farming data are collected from various sources to prepare data set), preprocessing of data, feature extraction, algorithm (KNN, Logistic Regression with PCA, and Hard Voting Classifier), recommendation system, recommended crop. Our dataset has around 1.5K data with eight parameters which are used in the dataset.

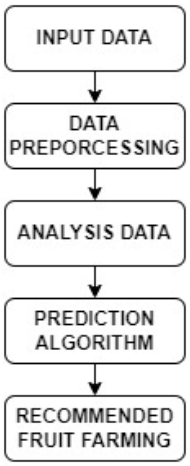

Fig 2 : Proposed Workflow Diagram

The parameter fruit farming is also used here which is used to predict as object data type. Soil parameter like Organic Carbon (OC), Potassium (K), Zinc(Zn), Sodium (Na), Nitrogen (N) & weather parameter like temperature and rainfall are used as float data type to predict the fruit. In our dataset fruit are used as target value.

V. METHODS

It is analytical method for estimating the current level given categorical data. To create proposed model Logistic Regression with PCA, KNN, ensemble algorithm and voting classifiers are used.

V.I Logistic Regression with Principle Component Analysis

Logistic Regression with PCA is a dimensionality reduction technique used in ML. It transformed high-dimensional data into lower-dimensional form while retaining as much of the original variability as possible. Logistic Regression with PCA works to find the principal components of data, which are new axes that capture the most variance in data from our dataset.

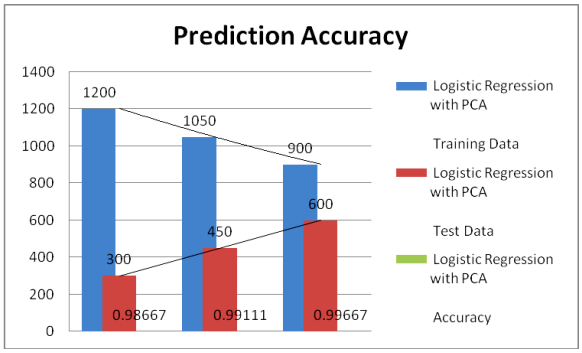

Fig 3: Accuracy analysis using Logistic Regression using PCA

Here 20%(300), 30%(450) & 40%(600) test data have taken respectively from our dataset (1500 tuples), & the accuracy using the said data are demonstrate using the Chart in Fig.3.

V.II K-Nearest Neighbor:

K-Nearest Neighbour is one of the simplest ML based on Supervised Learning technique. Based on similarity it classifies a new data point and store alldata fruit farming data from out dataset having around 1500 rows and eight attributes including target attribute.

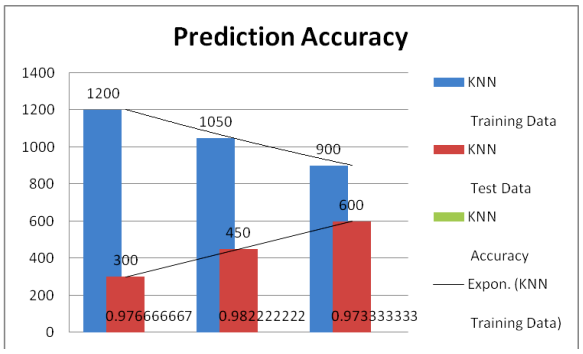

Fig 4: Accuracy analysis using KNN

Here 20%(300), 30%(450) & 40%(600) data have taken as test data respectively from our dataset (1500 tuples) & the accuracy, using the said data are demonstrate using the Chart in Fig.4.

V.III Voting Classifier (Hard Voting)

The majority voting [2] [4] is calculate from individual classifier votes. Each classifier makes a prediction, and the ensemble's prediction is simply the majority vote. Here algorithm works on each classifier to make fruit farming prediction. The ensemble prediction is done on major voting from each classifier of Regression using PCA& KNN.

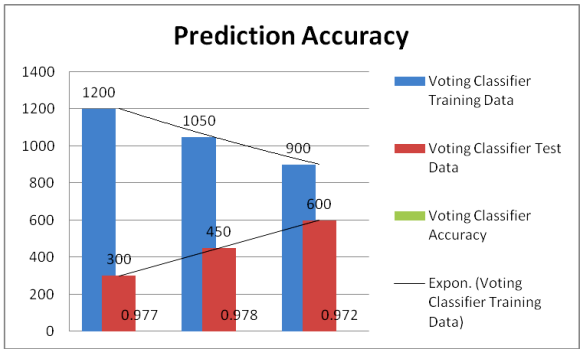

Fig 5: Accuracy analysis using Hard Voting Classifier

Here 20% (300), 30% (450) & 40% (600) test data respectively have taken from our dataset (1500 tuples) & the accuracy using the said data are demonstrate using the Chart in Fig.5.

VI. OUTPUT:

Accuracy level in Logistic Regression with PCA is better than other as the average accuracy is 0.991481481. The comparative of accuracy using are tabulated in table 1.

Table 1: Classifier Accuracy Comparison Table

Prediction Accuracy Comparison Table				
Training Data	*Test Data*	*Accuracy using Logistic Regression with PCA*	*Accuracy using KNN*	*Accuracy using Hard Voting*
1200	300	0.98667	0.976667	0.977
1050	450	0.99111	0.982222	0.978
900	600	0.99667	0.973333	0.972
Average Accuracy		**0.991481481**	0.977407	0.975667

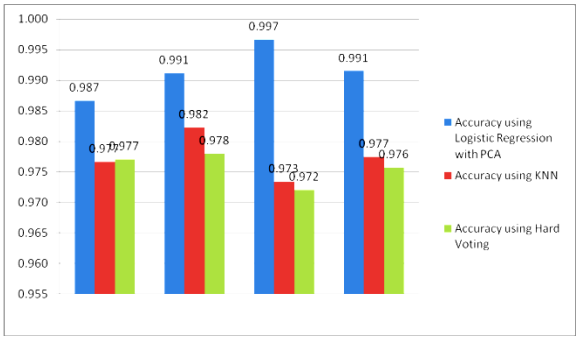

Fig 6: Accuracy comparison analysis Logistic Regression using PCA, KNN & Hard Voting Classifier

Using sampling data with KNN classifier predict the fruits more accurately by the proposed model.

Now the output data of recommended fruit farming of five different sets of sampling data which are taken randomly using KNN classifier from dataset.

Output of Random Sampling

```
['apricot']
     Organic Carbon  Potassium     Zn  Sodium  Temperature  Nitrogen  Rainfall
1359            6.11        6.0  85.87    55.0        24.89      81.0     51.71
```

Fig 7.1: Output-I of Random Sampling

```
['apricot']
     Organic Carbon  Potassium     Zn  Sodium  Temperature  Nitrogen  Rainfall
490            5.11       55.0  69.69    16.0        21.01      23.0    185.2
```

Fig 7.2: Output-II

```
['grapes']
     Organic Carbon  Potassium     Zn  Sodium  Temperature  Nitrogen  Rainfall
589            5.26       45.0  55.21    18.0        30.44      40.0     30.92
```

Fig 7.3: Output-III

```
['grapes']
     Organic Carbon  Potassium     Zn  Sodium  Temperature  Nitrogen  Rainfall
827            7.67       72.0  66.71    15.0        22.99       6.0     54.49
```

Fig 7.4: Output-IV

```
['apricot']
     Organic Carbon  Potassium     Zn  Sodium  Temperature  Nitrogen  Rainfall
17             6.97       35.0  80.42    39.0        23.79      91.0    206.26
```

Fig 7.5: Output-V

VII. CONCLUSION AND FUTURE WORK

This proposed recommendation systems informed about what fruits to grow based on various environmental factors and soil parameters. The classifier such as Logistic Regression with PCA, KNN and Hard Voting Classifier is used in the accurate and reliable fruit recommendation model.

Data preprocessing, model training, evaluation are important steps in building an accurate and reliable fruit recommendation model. These techniques can help to ensure that model is representative of the entire population and can make accurate fruit farming predictions for new and unseen data.

To conclude, this paper focuses on fruit productivity based on ML for prediction and recommendation. For prediction, Logistic Regression with PCA, KNN, and Hard Voting Classifier have been applied. Here Hard Voting Classifier shows the better result than other algorithm.

The future work will include the web application from where user can get the accurate suggestion for fruit farming by providing local soil parameter & relevant environmental parameters. That will integrate with the weather condition. Future fruit advisory systems must be developed to update climate models and provide climate change- sensitive recommendations such as natural calamities like droughts, floods and hot winds.

REFERENCE

[1] P. Devika,Suwarna Gothane; Degala Divya Priya, "Fruit disease detection using image processing and ML", AIP Conf. Proc.2492030064-023), https://doi.org/10.1063/5.0131092

[2] Khalied Albarrak,Yonis Gulzar,Yasir Hamid, Abid Mehmood, Arjumand Bano Soomro, "A Deep Learning-Based Model for Date Fruit Classification", 2022, 14(10), 6339; https://doi.org /10.3390 /su14106339

[3] J. Zhao, J. Tow and J. Katupitiya ,"On-tree fruit recognition using texture properties and color data," 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, AB, Canada, 2005, pp. 263-268, doi: 10.1109/IROS.2005.1545592.

[4] L. Hou, Q. Wu, Q. Sun, H. Yang and P. Li, "Fruit recognition based on convolution neural network," 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, 2016, pp. 18-22, doi: 10.1109/FSKD.2016.7603144.

[5] N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar and N. K. Cauvery, "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018, pp. 114-119, doi: 10.1109/CSITSS.2018.8768790