CROP RECOMMENDATION SYSTEM USING ML

Falguni Adhikary[1], Swarup Kumar Paul[2], Abhijit Das[3]
RCC Institute of Information Technology, Kolkata, India

Abstract- **The Indian economy is dominated by the agricultural sector. Agriculture is one the largest employment sectors in India. Most of the population of India is employed in the agricultural sector. Although India is an agriculture-based country, the Gross domestic product (GDP) of this sector is not enough. Only 16 percent is contributed to the agricultural sector.**

Due to industrialization, excessive use of pesticides badly affects soil productiveness. Most of the methods implemented by farmers using their experiences are not enough to increase productivity as needed. A common problem among Indian farmers is that they do not choose the best quality products due to non adequate knowledge & financial constraints. The goal of production is to maximize the yield and minimize the farming cost.

The crop recommendation process is to provide farmers with accurate advice based on their geographical region and demands. It uses a variety of input data such as soil characteristics considering parameters of N, K, P, pH, rainfall and temperature to recommend the appropriate crop. Traditional crop recommendation systems [1] do not take all the parameters that may affect product performance. The system integrates inputs that make predictions of crops in advance & the farmers to choose products for avoiding loss. The planning process can also help farmers select crops for rotation.

Algorithms such as machine learning i.e. Random Forest (RF), Support Vector Machine (SVM), Naive Bayes, ensemble machine learning can analyze large amounts of datasets to make more accurate decisions. Using advanced techniques it increases crop yields[2], improves resource management & greater profitability for farmers.

Index Terms- **Random Forest (RF), Support Vector Machine (SVM), Naive Bayes, ensemble machine learning, Gross domestic product (GDP)**

I. INTRODUCTION

The cultivated areas of India are very large, though it is not possible to achieve the maximum yield due to lack of rainfall, lack of water and farmers may not cultivate their crops in time. Indians have faced floods, earth-quakes, droughts, storms, water scarcity etc. As India's agriculture is diverse in nature, many crops are grown based on different regions and climates[3]. The government, researchers and other stakeholders are continuing working to meet these challenges and obtain continual development in India's agriculture sector.

Crops grown in India include wheat, rice, jute, bananas, potatoes, black gram, mungbean, black gram, tea, papaya, coconut, cotton etc. The main production of rice & wheat is done in India. The initiative has already been taken to expand the country's agriculture, including providing fertilizer, water subsidies, and promoting crop diversification. The government has also introduced programs like e-NAM (National Agricultural Market) to increase the production and increase income of the farmers. Digital India is also encouraging farmers to improve farming strategies by adopting new technologies. Using these new technologies, farmers can change their old practices, increase sustainable productivity & improve their living.

But farmers are cultivating the crops according to their past experience. Therefore, most of the time they face losses due to not adequate knowledge about what kind of crops to grow. Due to the lack of proper crop planning, farmers often commit suicide due to huge loss. Our model helps farmers to get a good idea of what crops to be planted based on their field. If technology is used in agriculture, farmers can know the needs of the product also. Crop recommendation systems have the potential to make advanced decisions regarding crop production, reducing huge loss and maximizing profits.

This proposed model also suggests the crop in advance in their field considering the parameters of Nitrogen (N), Potassium (K), Phosphorus (P), temperature, humidity, pH, and rainfall [4]. Using our model, a farmer can identify the profitable crop to grow on his land using the right combination & reduce the risk factor of loss. The government can also make the policies to which crops to be imported or exported.

II. PROBLEM STATEMENT

Collecting relevant and accurate information on soil quality, weather patterns and crop yields is critical to develop crop recommendations. However, data collection can be a major problem, especially in rural areas with no internet connectivity and infrastructure. Additionally, the data quality may vary, resulting in incorrect recommendations.

It requires the integration and analysis of many data, including soil characteristics, weather data and crop yields. Analyzing and processing this information requires complex processes and computing power is difficult for farmers without the skills or technology.

Weather conditions, pests and other environmental factors can affect crop yield. Plant recommendations should take these changes into account to make recommendations. However, accurately predicting these conditions can be difficult, especially in areas with highly variable weather patterns.

Developing and implementing crop recommendations can be costly and limit their use by farmers in low-income countries. In addition, providing farmers with training and support in the correct use of the system can be difficult, especially in areas with low literacy or technology use.

Advanced crop recommendation systems provide the best ideas of crops in their fields considering various factors N, K, P, pH, temperature and rainfall.

This project will analyze agricultural data such as land availability, weather conditions and recommend the best crops. The challenge behind the new proposal is to collect agricultural data, which must be accurate, reliable and up-to-date as we recommend the right crops.

III. LITERATURE STUDY

The proposed planting system is designed to provide farmers with information about the crops to grow in an area based on various factors such as soil type, climate and other environmental factors. In recent years, due to the importance of precision agriculture, information on recommended planting has increased rapidly.

Nidhi H Kulkarni, et al. used to create models that combine the predictions of multiple machine learning models to recommend the right crop on the specific type and characteristics of the soil. The independent learners used in mixed models are RF, Naive Bayes and Linear Support Vector Machines. Distribute essays written by one of the environmentalists to the class using the majority voting system. Recommended crops are divided by soil information entered into Kharif and Rabi recommended crops.

The data includes climate factors such as average precipitation and temperature patterns, as well as specific soil physical and chemical properties. A combination of independent learners achieved a classification accuracy of 99.91%.

D. Anantha Reddy, et al. offers solutions such as suggestion methods from mixed models where most voting methods use random trees, CHAID, K_Nearest Neighbors and Naive Bayes as candidates. This approach offers solutions such as suggestion methods from mixed models where most voting methods use random trees, CHAID, K_Nearest Neighbors and Naive Bayes as candidates. The discrete image generated by this strategy includes analysis of actual soil conditions, crop yield, state and regional crop quality to predict crop yields, particularly in specific climates.

Javier Lacasta, et al. process aims to help identify pests and select appropriate treatments. The core of the process is an ontology that models interactions between crops, pests, and treatment. To use this information, these projects have created a consensus to help identify pests and select appropriate treatment. At the heart of the system is an ontology that models interactions between crops, pests and treatments.

K. Anji Reddy, et al. uses collaborative filtering, which provides some recommendations to users based on a comparison of user behavior and operating patterns, and suggests preferences and behaviors as well as users. It also tries to predict the suitability of a product for a particular process. Such an optimal system can show crops that can grow according to the soil and climate. The research focuses on developing a consensus system that can collect raw data on environmental factors such as soil and air quality from experienced farmers, economist farmers and other stakeholders.

Michael Neugart, et al. analyzed showing that the elite received 2.2 percent (or 0.9 percent) of the vote in flooded cities in the east of the country, along with the west. Analysis of the early floods, changes in East Germany's democratic knowledge and panel studies provide further evidence that voters less aware of democracy are more likely to lose their right to vote.

IV. PROPOSED SYSTEM

Our model consisting of six stages that are demonstrated in below block diagram & that are dataset preparation (agricultural data are collected from various sources to prepare data set), preprocessing of data, feature extraction, algorithm (RF, SVM, Navie Bayes, Ensemble Algo), recommendation system, recommended crop. Our dataset has more than 2K data with eight parameters which are used in the dataset.

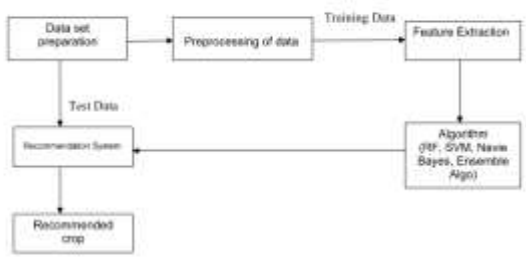

Fig 2.1: Block diagram of proposed framework

The parameter crop is also used here which is to be predicted as object data type. Crop represents different types of crops specially cultivated in India. Here Nitrogen is int type & remaining six parameters are float data type.In our dataset crops are like rice, maize, chickpea, kidney beans, pigeonpeas, moth beans, mungbean, black gram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee.

V. METHODS

This training data is fed into the system to teach the machine how it works. Similar to humans, AI training data is similar to the instructions and experiences we encounter as we learn, process data, and make decisions. Because ML data is essential for prediction and efficiency needs to be under human control from beginning to end of process training. If an algorithm is well trained, it successfully identifies features, makes connections with data, and increases performance.

A sample of data from a training model that is often used to predict the prediction of a model's ability when adjusting the model's hyper parameters.

Actual test data are different from test data that are not included in the training model but used to provide an unbiased assessment of the skills of the final adjusted model for selection or sample comparison.

For fit the model training data is used. The proof of validation is data set used to provide an objective assessment of how model fits the data subject to the hyper parameters. Evaluations are increasing ability to analyze data is important part of setting a standard.

Test data is used for evaluation. There are other ways to calculate biased estimates or to increase model skill estimates on data that is not visible in context of current data.

In ML data preprocessing, we divide our dataset into two sets. This is a very crucial step of data pre-processing as by doing this we can enhance the performances of our ML model. The training dataset is used to train the model & develop the model. Test data set can be used after training is done. Training data can differ on center of supervised [5] & unsupervised algorithms.

Once our machine learning model has been trained with 80%, 70%, 60%, 50% of data from our dataset as training data & 20%, 30%, 40%, 50% of data as test data respectively. In this phase we also can check the accuracy of our proposed system.

A. Random Forest Classifier

Random Forest Classifier [6] can be useful tools in the recommendation process because they can help farmers decide which crops to plant in which area, especially based on the unique environment of agriculture.

Percentage of test data & train data are taken from our dataset as 20%, 30%,40%,50%,60% test data and remaining 80%, 70%,60%,50%,40% training data respectively.

The accuracy percentage of RF is as follows:

Test Data	Training Data	Accuracy
440	1760	76.60
660	1540	77.42
880	1320	74.31

1100	1100	77.18
1320	880	76.59

Table 5.1: Accuracy table of RF

B. Naive Bayes

The classifier can learn about various crop conditions and predicts which crops are best suited for the process.

Percentage of test data & train data are taken from our dataset as 20%, 30%,40%,50%,60% test data and remaining 80%, 70%,60%,50%,40% training data respectively.

The accuracy percentage of RF is as follows:

Test Data	Training Data	Accuracy
440	1760	75.45
660	1540	75.45
880	1320	74.77
1100	1100	75.54
1320	880	75.83

Table 5.2: Accuracy table of Naive Bayes algorithm

C. Support Vector Machine

Support Vector Machine [7] is concerned with binary data classification. Basis on seven parameters, SVMs determines whether crops can grow in that area. This method creates a large plane in N-dimensional space, where N is the number of features that will be used to describe the content of the data in the dataset. The greater the distance of two points, the more accurate the distribution.

Test Data	Training Data	Accuracy
440	1760	75.90
660	1540	74.09
880	1320	72.84
1100	1100	74.09
1320	880	75.90

Table. 5.3: Accuracy table of SVM algorithm

D. Voting Classifier

Voting is an integrated method of ML/regression, a bunch of votes that involve making an estimate that is mean of many other regression models. In the distribution, the final voting group[8] will include voting for the clean class list from other models and prediction class with most votes. Soft polling involves estimating the probability class list and predicting the list with highest probability.

Random Forest	SVM
97.27	97.27
91.17	92.20
80.30	83.64
80.36	83.27
70.90	75.0

Table 5.4: Result of accuracy using voting classifier

VI. OUTPUT:

The input parameter & corresponding values are listed in below mentioned tabular format.

Input Parameter	Value
Nitrogen	42
Phosphorus	65.65
Potassium	75.895
temperature	18.74
humidity	18.11
ph	8.33
Rainfall	59.52

The recommended crop:

```
['jute']
     Nitrogen  Phosphorus  Potassium  temperature  humidity    ph  rainfall
208        42       65.65     75.895        18.74     18.11  8.33     59.52
```

Here the parameter of Nitrogen 42, Phosphorus 65.65, Potassium 79.895, temperature 18.74, humidity 18.11, ph 8.33 & rainfall 59.52 is taken randomly from our dataset. The predicted crop is jute here.

```
['mango']
      Nitrogen  Phosphorus  Potassium  temperature  humidity    ph  rainfall
1185        39       19.65     27.895         36.1     52.16  6.25     98.65
```

Again using sampling data from the dataset it is showing mango. According our data set it is predicted as true.

VII. CONCLUSION AND FUTURE WORK

Algorithms for crop recommendation systems informed about what crops to grow based on various environmental factors and soil conditions. Ensemble learning techniques such as Naive Bayes, Support

Vector Machines, and RF is used for accurate and reliable crop recommendation models.

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

REFERENCES

[1] S.kanaga Subba Raju et al, Demand based crop recommender system for farmers, International Conference on Technological Innovations in ICT For Agriculture and Rural Development, 2017.

[2] Shriya Sahu et al," An Efficient Analysis of Crop Yield Prediction Using Hadoop framework based on RF approach, International Conference on Computing, Communication and Automation, 2017.

[3] Ranney, J. W., and L. K. Mann. "Environmental considerations in energy crop production." *Biomass and Bioenergy* 6.3 (1994): 211-228.

[4] S. Bhanumathi, M. Vineeth and N. Rohit, "Crop Yield Prediction and Efficient use of Fertilizers," *2019 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, 2019, pp. 0769-0773, doi: 10.1109/ICCSP.2019.8698087.

[5] Suganya, M. "Crop yield prediction using supervised learning techniques." *International Journal of Computer Engineering and Technology* 11.2 (2020).

[6] Geetha, V., et al. "An effective crop prediction using random forest algorithm." *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*. IEEE, 2020.

[7] Gandhi, Niketa, et al. "Rice crop yield prediction in India using support vector machines." *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE, 2016.

[8] Archana, K., and K. G. Saranya. "Crop yield prediction, forecasting and fertilizer recommendation using voting based ensemble classifier." *SSRG Int. J. Comput. Sci. Eng* 7 (2020): 1-4.