Enhancing Crops Production Based on Environmental Status Using Machine Learning Techniques

Shiyam Talukder (talukdershiyam@gmail.com)
Habiba Jannat (habiba.jannat@gmail.com)
Sukanta Saha (sukantasaha0182.ss@gmail.com)
Katha Sengupta (kathasengupta11@gmail.com)
Dept. of Computer Science and Engineering, BRAC University
Dhaka,Bangladesh

***Abstract*—Bangladesh is an agricultural country. As the economy is based on agriculture highly, there should be progress in this sector. To make progress in agriculture the productivity must be increased. These days, productivity is low due to various factors. One of them is not finding suitable crops for a particular land. In this way, the crops are not produced at the maximum amount. Hence, productivity of agriculture depends on multiple parameters on the basis of location. The suitable crop for a particular location is necessary for agriculture to bring the most productivity. Here we have designed a model that predicts productivity with given parameters, and also recommends the suitable crop based on those parameters. In terms of Machine Learning for the prediction and the recommendation, we have applied multiple algorithms like k-nearest neighbor, support vector machines, random forest, naive bayes' classifier and logistic regression, collaborative filtering and fuzzy K-Nearest neighbor. After training the dataset and applying algorithms, for prediction we have made a comparison by analyzing the precision. On the other hand, for recommendation we have used collaborative filtering system and fuzzy k-nearest neighbor. These algorithms are mainly used to take users data as input and test with the trained data that is already in the system and will filter out the best 5 crops as output.**

***Index Terms*—Agricultural Productivity; Recommendation; Prediction; Precision; K-Nearest Neighbor; Collaborative Filtering; Fuzzy K-Nearest Neighbor**

I. Introduction

EVER since the very beginning of human civilization, agriculture has been the root of the development. As food is one of the basic human needs, agriculture has been the source of the supply of it. On the other hand, agriculture affects the economy by providing raw material and food to non-agricultural sectors. It has a good support of employment amongst rural dwellers, providing them with purchasing power. The dominance of agriculture can be observed from non-developing countries to developed countries. The developed countries have a history of good agricultural productivity that led to the enhancement of the economy. The agricultural activities give the economy some power so that it can enrich itself for betterment and upgrades the economy. Thus, agriculture has given a proper structure to the source of investment. As for developing countries like Bangladesh, most of the industries here are based on agricultural products including pharmaceutical and food processing. In this way, it has always played a strong backbone for a nation's flourishment. Back in the Neolithic revolution, agriculture empowered the growth of the human population more than hunting or gathering. Thus, agriculture is the root of human survival.

Even with the transformation of the agrarian societies into industrial societies, as it mentioned before, countries like Bangladesh depend on agriculture as it plays a vital role on macro-economic objectives like reducing unemployment or poverty, achieving food security. The economy here has an increased demand of other raw materials that comes from agriculture to support the population. But the drastic increase of population with the increasing demand of staple food and the decreasing arable land has a negative effect on the economy. The agricultural GDP is declining day by day and it clearly shows how we are not getting enough efficiency to produce the crops. The efficiency that can be measured by making choices by the selection of suitable crops for suitable lands.

Since the foundation of agrarianism we have noticed how much agriculture is dependent on multiple parameters of the environment and surroundings. Even though agriculture has been reducing hunger and other basic human needs, the sector of agriculture has been underdeveloped. There is a lot of failure, lack of production that takes place for the need of making rational choices. If there has been rational choice in agriculture, then the productivity would be high and there would be more profit from it as the proper choices of agriculture would reduce the cost and money. Hence, there should be a proper study for the betterment and benefits in it so that cultivators get more results than the effort. The failure is very unexpected, the investment in agriculture is very crucial and struggling for developing countries. If there are predictions that cannot take place, then there can be more losses than profits. As agriculture depends on nature, the growth of the crops depends on various components of the climate such as rainfall, humidity, temperature and also it is dependent on multiple seasons. Thus the determination of choosing the suitable crop amongst the alternatives for a particular area can be very challenging to decide by the cultivators. It is high time to achieve maximum production of the crops to reduce hunger

and poverty as agriculture deals with food security and as well as growth in the economy.

II. Literature Review

Here we have studied the application of different machine learning algorithms in various ways. For instance, In [1], D. S. Zingade et al. used various machine learning techniques to build a system where it takes the data of weather and soil condition and based on it the system provides a solution to predict the most profitable crop for. The system is an Android based application that analyses the data for accurate output. The system combines the data from the repository and weather department and implements machine learning algorithms. Here the algorithm they have applied is multiple linear regression which is a prediction algorithm. The system takes the input by the help of GPS and by that the system provides the profitable crop to the farmer. The previous data is also in the consideration of the system for the most profitable crop to the desired area. In this way the result is found through training and testing and the algorithms consider the input of weather, soil. After this, data process the cultivable crops with the help of previous production and predict through multi-linear regression with weather, soil material and previous production then the system calculate the profit through the parameter of prediction and current crop and it shows the output of maximum profitable crops with proper requirements which works for suggestion. In this way the system works with agricultural productivity helping the cultivators take the decision of proper crop and also gives the suggestion to the them. the system takes the idea of previous production, soil, area into consideration for the best result.

Not only we studied about the algorithms but also we observed the variation of databases. Such as, in [2], S. S. Dahikar and S. V. Rode implemented the ANN algorithm to determine the suitable crop for the specific soil type. As the soil has a lot of parameters such as pH, nitrogen, temperature and the effect of the climate on the soil, they have predicted one particular crop to be yielded in one particular type of soil depending on the parameters. Here they had a comparison of different types of algorithms such as artificial neural network, information fuzzy network and other data mining techniques and with the comparison they have come to a conclusion where they have claimed the Artificial Neural network algorithm is the most suitable to their system. This paper provides a comparison study of different algorithms where there is accuracy measured and the best algorithm is chosen. They have claimed in the algorithm that the Artificial Neural network is good for prediction to their system using the soil, weather and market price as the given parameter.

The accuracy of some algorithms has performed better in some situations. Such as, in [3], Mr. V. Lamba and Dr. V. S. Dhaka discussed the techniques of wheat yield prediction comparing Artificial Neural Network (ANN) with other models such as Multiple Linear Regression (MLR), Logistic regression, Time series model etc. With their research they have claimed the accuracy and efficiency of ANN better than others. They have used different types of models on different aspects. For Multiple Linear regression and logistic regression models they predicted crop yielding through plant characters and pest count. They have implemented the probabilistic model Markov chain to forecast the techniques in agriculture. Moreover, In [4], authors have compared between different AI models for production in the Midwestern region crops in the United States. They have initially selected the summertime for specific crops and applied the model on it. With a comprehensive study they had a comparison and with proper measurements they have proven that the Deep Neural Network is suitable for their study. They have used satellite images, meteorological data, hydrological data, crop yield statistics and cropland for the dataset in such a way so that it can give the desired outcome. The satellite products have been used to extract cropland pixels for corn. In this way they have combined different techniques for an ideal solution. To summarize, Using the DNN model they have performed an optimization process for the accuracy and they have found it has outperformed rest five of them with prediction error around 7.6% for corn and 7.8% for soya bean for which the error is much less than other methods. In [5], the authors developed a model for forecasting agricultural products by a new type of model which is able to perform on both linear and nonlinear forecasting functions. They collected export data of shrimp and chicken of Thailand performed the algorithms on it. Here they used a model combined with several mathematical equations and then SVM was applied and cross- validation process was performed. Their study showed that SVM with cross-validation is accurate forecasting tools than others for their dataset.

For the recommendation part we have also studied the techniques to apply by relating with other problems. For instance, it has always been the part of preferences in other words the recommendation engines are built by similar choices or similar types of users. Hence, we have gotten our ideas and approaches to apply to work on our problems for the recommendation through the related papers. In [6] the authors have developed a framework for E-learning recommender. Here they have proposed a platform where the system has multiple layers including recommendation models and supplementary data like web logs. They have discussed multiple algorithms that can be implemented to build recommendation engines such as content based, collaborative bas r or knowledge based, hybrid filtering recommendation. From this paper we have learned about a few algorithms, especially for knowledge based where the user ratings are insufficient, or hybrid filtering the recommender combines the advantages of several techniques to minimize the problems. Moreover, we also have related our problem with webpage viewing. Such as, in [7] authors have built a recommender system called PRES with object-oriented programming language Java. It has implemented the content-based filtering by vector space model where documents and profiles have been represented as vectors. Based on the profiles and recently visited pages, the documents have been matched. Hence, the user matches with the preferred document through previous interactions and preferences. This system has shown how users find their preferred pages with the algorithm implied.

Several recommendation systems have been implied in the

field of agriculture. In [8], authors have applied collaborative filtering that offers recommendation to the users by developing a system with the help of big data for the complexity. They have clustered the data by the characteristic similarities using an agglomerative hierarchical clustering algorithm, then applied Pearson correlation coefficient for the similarity (Collaborative filtering). With this system, farmers buy and sell the products by the application they have designed. Also, in [9], M. Kuanr, B. K. Rath and S. N. Mohanty built a crop recommender system with inference engine and knowledge-based engine together with input variables as temperature, humidity and rainfall. They have collected the data based on questionnaire assessment then created a database to predict the outcome of suitable crop. They applied cosine similarity to find similar farmers then used the inference rules for the recommendation. They have also proposed pesticides, fertilizers and other seeds.

III. Algorithms

A. Algorithms for Precision

1) Naive Bayes Classifier: It is a supervised learning algorithm which is used to classify data into predefined classes. It is a conditional independent because [10] all attributes are independent given the value of the class variable. Its main uses are in text classification, content arrangement, spam separating, sentiment analysis and many more. It is fast and easy to work with data. Therefore, it is a method for predicting the probability of independent attributes given a class using the Bayes' Theorem. The Bayes' Theorem:

$$P(A|B) = P(B|A)P(A)/P(B) \tag{1}$$

Where P(A—B) is the probability that is needed to find the naive classifier
Ultimately, for prediction the formula used is

$$y = argmaxp(y = C_k) \prod p(x|y = C_k) \tag{2}$$

2) Logistic Regression: It is the statistical model which is used to estimate the probability of the categorical values. It is a classification method for which it gives the output as binary value which is 0 (yes) and 1(no).It has a threshold value from which it predicts to put 1 or 0 where it is appropriate i.e $0 \leq t \geq 1$ It finds the probability of [11]the relationship between one dependent variable and one independent variable using logistic function. [12]The logistic curve is called sigmoid curve which is like S-shaped curve and it is used to find the sigmoid functional probability by

$$p = 1/1 + e^{-y} \tag{3}$$

where y = mx + c which does not give a linear curve. Sigmoid Probability on linear regression equation

$$p = 1/1 + e^{-\beta 0 + \beta X1 + ... + \beta nXn} \tag{4}$$

3) K-Nearest Neighbor (KNN): It is a supervised learning model and can be used for both classification and regression process. [13]KNN is a lazy learning and non-parametric algorithm because it does not need any training data rather it can take data directly from the dataset and use it in testing for making a model. Although it is slower and time consuming in testing, it is an easy and simple algorithm for finding the k closest value of the dataset. In KNN algorithm, first data from the dataset is taken to put it in training and testing where K value means the nearest or closest value will be selected. In order to select the k-closest value, the distance between points needs to be calculated. This can be done by different calculating methods such as Euclidean, Manhattan and Hamming distance where Euclidean distance calculating method is used frequently. After calculating the distance, arranging the data points in ascending order and then selecting the k nearest value and classifying the points. For reducing the error, repeat the process again and again and choose different k nearest values while maintaining the accurate predictions.

$$Euclidean = \sqrt{\sum_{i=1}^{k} \Big(x_i - y_i\Big)^2} \tag{5}$$

4) Support Vector Machine (SVM): SVM is a supervised machine-learning algorithm and it is a structural risk minimization which minimizes expected error of a learning model and for classification. [14] SVM uses kernel trick for non-linear Separation problems. [14] SVM is used to find a MMH and separates the data into classes. It separates the two or more data by measuring the distance of the margin which means the two distant lines of nearest points and making a hyperplane line between the maximized margin lines of support vectors of the data points. This hyperplane is also called decision boundary. [14]The hyperplane can be made by segregating the classes in the best way and selecting the right hyperplane with the maximum segregation from the either closest data points.

5) Random Forest: Random Forest is a supervised learning algorithm. It is based on the concept of Ensemble learning. Random Forest uses the dataset to take the data by feature bagging which is randomly selecting data into different sets of the training data and from that data different decision trees will be generated and prediction will be on the majority of the vote. Repeatedly the method is running for better accuracy of the prediction. Therefore, the random forest method introduces more randomness and diversity. It also lets us determine how much the model is overfitting.

B. Algorithms for Recommendation

1) Fuzzy KNN: Fuzzy KNN is the improved version of the most famous supervised algorithm called KNN and this is because KNN is a simple, effective and easy classification method with no preprocessing requirement. Despite the useful facts of KNN, it has some flaws, that is if the class is unbalanced, it cannot give better performance. It means the data in the training set is not properly accurate and it may decrease the precision of classification. Hence, the theory of fKNN is constructed with improvement in order to have less

error in classification. Fuzzy KNN algorithm [15]provides a more realistic vector of membership for the objects and accounts for the degree of object membership to the classes of objects. Basically, [15]class is assigned to the most common class. [15] fKNN algorithm is used mainly for dividing the cluster. The theory of Warren and Damin are used to put membership into different categories.

$$u_i(x) = \frac{\sum_{j=1}^{K} u_{ij}\left(1/ \parallel x - x_j \parallel^{2/(m-1)}\right)}{\sum_{j=1}^{K}\left(1/ \parallel x - x_j \parallel^{2/(m-1)}\right)} \quad (6)$$

[16]Where i = 1, 2, . . . ,c and j = 1, 2, . . . k, c is the number of classes, k is the number of k-nearest neighbours and m is the strength parameter. [16]Euclidean distance is between x and its jth closest points. uij is the degree of membership.

2) Collaborative-Filtering Algorithm: Collaborative filtering is a popular recommendation algorithm that uses people's opinion to filtrate or evaluate the process. It is based on the idea that the system will predict a result on how many people have rated a similar item and it will be repeated or suggested again in future. For example if u think of a movie rating system then CF feedback falls into two options. Firstly, "Explicit" is when users specify how much they liked a particular item by providing a numerical rating. Secondly, "Implicit" when a user chooses a movie that he likes to watch for that moment, the system deduces that the user is interested in that particular kind of genre of movie. CF algorithm has many forms and they are user-based and item-based collaborative filtering. CF has many uses in our daily life and those are Amazon, Netflix, iTunes, IMDB and many recommendation or predicted systems.In our model, item based collaborative filtering is used for the purpose of agricultural productivity. It is because item based CF method filters items that are suitable for that particular parameters using the matrix factorization process and similarity calculation.

$$simil(x,y)=cos(x.y)=\frac{x.y}{\|x\| \times \|y\|}=\frac{\sum_{i \in I_{xy}} r_{x,i} r_{y,i}}{\sqrt{\sum_{i \in I_x} r_{x,i}^2}\sqrt{\sum_{i \in I_y} r_{y,i}^2}} \quad (7)$$

IV. Proposed Model

A. Machine Learning Methodology

The workflow of our approach is given in Figure 3.1 to predict the productivity of crops and recommend the best suitable crops according to the Climate feature like humidity, rainfall and temperature. We have taken many approaches to make our prediction and recommendation more effective, accurate and efficient. During this approach we have pre-processed our data, select the best feature, train our model and test them with test data for accuracy and then we predict and recommend best possible crops. After applying those prediction and recommendation we have used may charts for visualizing our data.

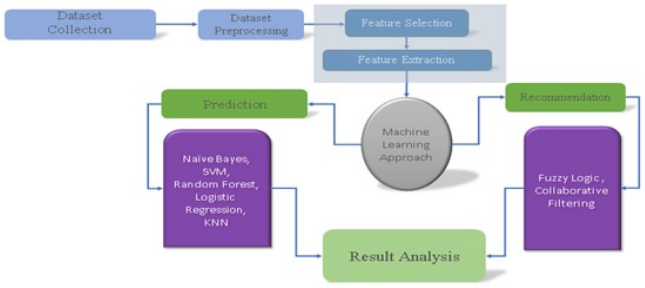

Fig. 1: Proposed Model

B. Dataset

Machine learning approach is a data-centric approach. Therefore, dataset plays a vital role in machine learning. We need to train our model to perform various actions by feeding data to our model. Our dataset is divided into two parts. One is training a dataset and another one is testing Dataset. Using training dataset, we train our model with various algorithms and we validate our model using testing dataset and get the necessary result.

C. Dataset Description

We have taken our dataset from an Indian Agriculture website by email request. In this dataset we have about 7 years of data record which contained 135 of different crops based on various categories. In the dataset we have 12 columns and total 11690 rows. The column names are 'Statename', 'districtname', 'year', 'season', 'area', 'rainfall', 'humidity', 'temperature', 'prevYearsRainfall', 'prevYearsTemperature', 'prevYearsHumidity', and 'Crops'. For back-end engines we used Keras and TensorFlow. A heatmap of the dataset is given in the Figure 2.

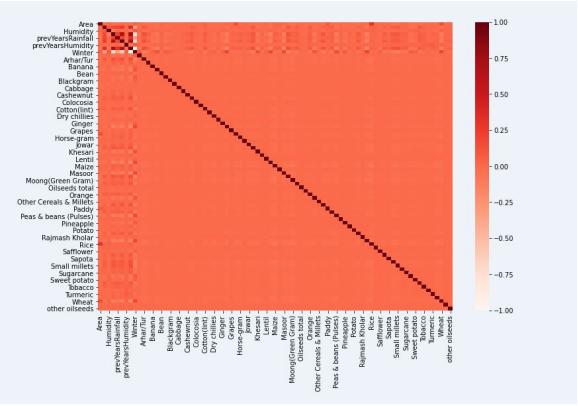

Fig. 2: Heatmap

A distribution plot of the most often crops like 'Rice', 'Maize', 'Moong', 'Arhar', 'Urad', and 'Wheat' vs different parameters are shown in the Figure 3, Figure 4, Figure 5 :

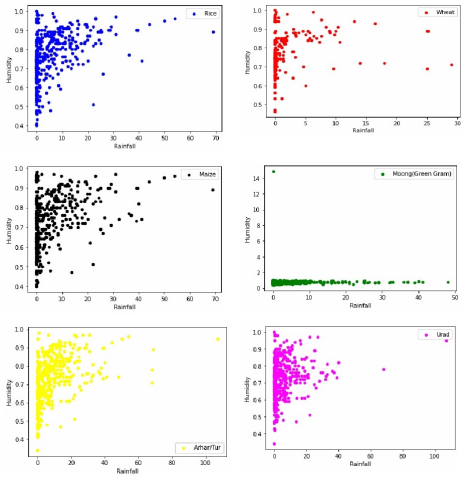

Fig. 3: Crops vs Humidity and Rainfall

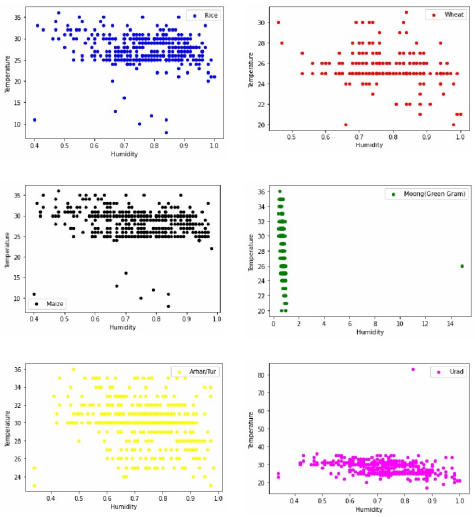

Fig. 4: Crops vs Humidity and Temperature

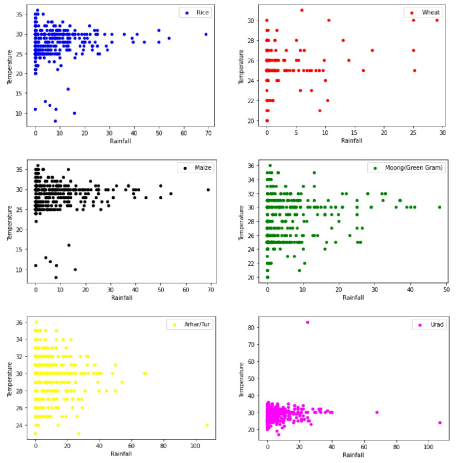

Fig. 5: Crops vs Rainfall and Temperature

D. Data Pre-processing

Data Pre-processing is a process where any data is transferred to machine readable form. Machines are unable to interpret real world data. Therefore, we have to process our data such a way that it is understandable by machines.

Firstly, to process our data we need to import the necessary library. Libraries are the collection of functions which can be called and used when necessary. Using this library, we can import our dataset and apply many functions according to our necessity.

Secondly, we can import our dataset using the library. Generally, the dataset is in a csv form. We can read that csv data using panda's library. Using this library, we can load our data into our machine.

Next, we need to reduce the unnecessary data from our dataset. Then, we have to check if there is any null value. If they exist then we can replace them by mean value. Again, there is lots of data which cannot be interpreted by machine. There-fore, if we want to work with them, we need to convert them into binary value.

Moreover, there are some categorical data which are tough for a machine to under-stand. Therefore, we need to encode the categorical data if we want to use that data into our model.

Now, we will divide our dataset into two parts, one is a training set and another one is testing set. We will build and try to train our machine learning models on our training set, i.e our machine learning models will try to find if there is any correlations between our training set and then we will test the models on our test set to check how accurately our model can predict the dependent variable. Generally, 80 percent data are used for testing and 20 percent data for training.

Now we need to divide our data into 2 sets, training and testing. Again, We will divide them into 4 sets—(1) X train (training part of independent variable), (2)X test (test part of the independent variable), (3) Y train (training part of the dependent variables based on the X train , Y test (test part of the dependent variables based on X test).We will split our data by 80 and 20 ratio or according to our algorithm choice

Finally, we need to scale our features. In Scaling we standardize our data ratio wise. In our model we need to use cosine similarity. Therefore, any negative value will give unreal value. We need to standardize the data to avoid negative value. To do this work, we will import the class Standard Scaler library from the scikit pre-processing library and use them when necessary. Now we will transform and fit our X train set. This will bring all the data into the same standard.

E. Feature Selection

Feature selection is the process of reducing the number of input variables when developing a predictive model. There are some model which have a lot of variable which can bring down the progress of our model. Again, data model which have an unnecessary variable can be a wastage of memory and speed. Moreover, Irrelevant input variable can degrade our model performance. In our dataset, State name, District Name, Crop

year, Previous year humidity, Previous year rainfall, previous year temperature does not have any impact in our model. Again, we have also removed all the Null values using the feature selector.

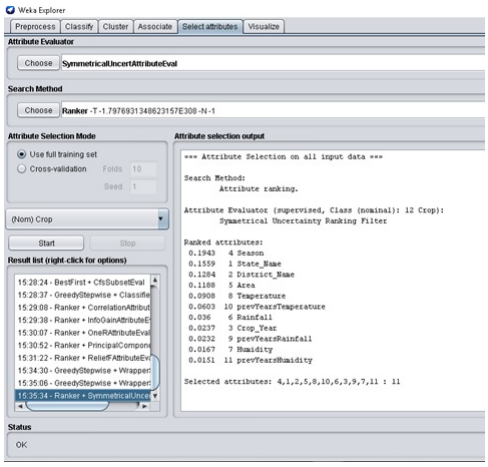

Fig. 6: Feature Selection using WeKa Explorer

For Selecting feature, we, have used a software called Weka Visuals. It is a software used for data mining. We can apply many algorithms in our data. It contains many tools so that we can pre-process our data. Weka is not only used for data preprocessing but also it can be used for classification, regression, visualization and data associative. We can use this software according to our need.
We have used "Symmetrical UncertAttribute Evaluation" which Evaluates the necessity of an attribute by measuring the symmetrical uncertainty with respect to the class. Ranker Ranks attributes by their individual evaluations. For selecting an attribute, we have used a full training dataset split into 80:20. However, we have used Season, Rainfall, Temperature, Humidity as a key attribute for our model.

F. Training and Testing of Machine Learning Classifier and Recommender

After we are done with our feature selection process, we will train and test our ma-chine learning classifier. Then we will divide our dataset into 2 set which is training and testing sets. Next, we train and test the dataset using many classification algorithms. Training data are used to train and build the models. Test data are used for understand how well the model is performing. Test data is used to predict answers dependent on training. In our dataset, we used eighty percent data as training data and twenty percent as testing data in prediction and for the recommendation we used 50 percent data for training and 50 percent data for testing. After splitting the dataset, we train our model. With various algorithms which are given Below.
For Prediction:
1. naïve Bayes classifier.
2. K-nearest neighbor
3. Random Forest
4. SVM
5. Logistic regression.
This prediction algorithm tries to predict the crop based on our selected features
For Recommendation
1. Fuzzy Logic
2. Collaborative Filtering
These recommendation algorithms tries to recommend some best possible crops based on their selected features like the rainfall, humidity and temperature.

V. Proposed Prediction Model

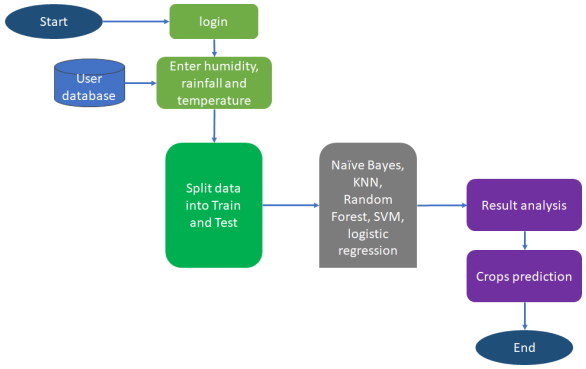

Fig. 7: Proposed Prediction Model

When a user gives the different weather parameter values in the dataset it will store the data in the dataset and run these following 5 algorithms by splitting the whole data into 80% of the train and 20% of the test. After that, it will evaluate the result and show the best crops predicted.

VI. Proposed Recommendation Model

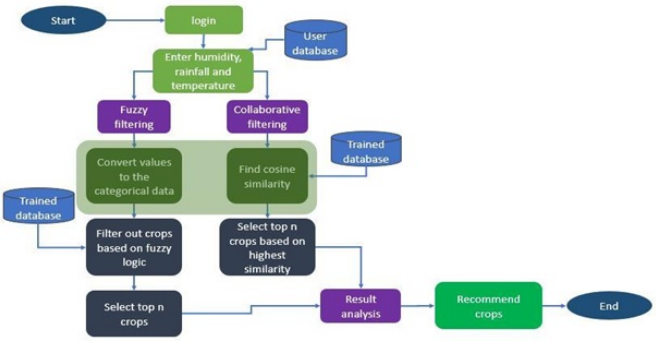

Fig. 8: Proposed Recommendation Model

In our recommendation model we have implemented two algorithms: Collaborative Filtering and Fuzzy K-nearest neighbor.
In Collaborative Filtering, Use will give their humidity, temperature, rainfall values. Then these values will be sent to our server. In Server we have our trained data. It will check the similarity between our trained data and input data.

Then, it will filter out the 5 based crops based on their cosine similarity. Then, the input value with current crops will be updated in our model. Thus, it will improve the accuracy of our model day by day.
In Fuzzy K-nearest neighbor, User will give three inputs of humidity, rainfall and temperature. Based on the value of this parameter the value will set under three categories 'High', 'Low', 'Moderate'. Then this data will be sent to our server. In our server, we have categorized our crops based on 27 conditions. When any of these 27-conditions match with our input, it will show the top 5 crops which are suitable based on a given condition. Our server will be updated continuously therefore; user will have a vast choice of crops based on their climatic condition.

VII. Result Analysis

A. Result Prediction on Algorithms

We implement five prediction algorithms and two recommendation algorithms, we also analyzed their precision.

Algorithms	Precision	recall	f1-score
Naïve Bayse	72	60	58
K-Nearest Neighbor(KNN)	64	60	61
Random Forest	70	69	70
Support Vector Machine (SVM)	56	61	53
Logistic Regression	70	67	68

TABLE I: Algorithm Precision

We got the highest precision in Naïve Bayse that around 72. The lowest precision we got is 56 in SVM. For KNN we got precision of 64. Both Random forest and Logistic Regression we got precision of around 70. A comparison of the precision of these algorithms are given in figure 9.

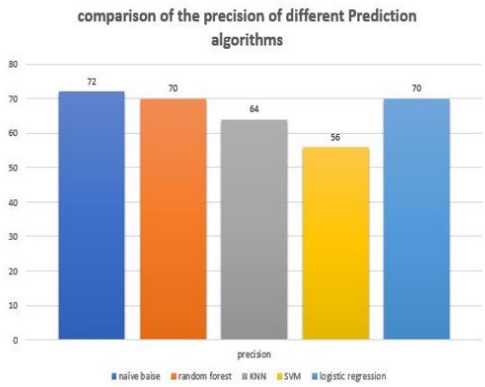

Fig. 9: Precision comparison of different Prediction Algorithms

B. Predicted Crops

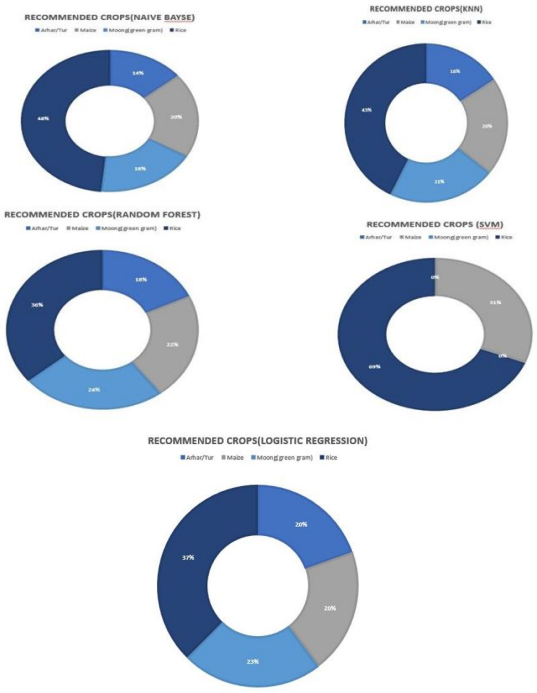

Fig. 10: Precision of the Top 4 Crops

We took the top 4 crops that have the highest precision in the following prediction algorithms and made a pie chat out of it. Here, Rice has the highest precision in every algorithm compared to other crops like Moong, Maize and Arhar/Tur. It is 48% in naive bayes', 43% in knn, 36% in random forest, 37% in logistic regression and greater portion of all in svm with 69%.

C. Collaborative filtering

Here, In our dataset in our dataset Rice were cultivated in 956 places, Bajra in 189 places, Urad in 472 places, Groundnut in 241 places, Maize in 900, Arhar in 478 places, Mung in 666 places. When we use cosine similarity to find similar geographical attribute then this crop had the best similar value. Using the similarity, we find that Maize has 2280 places with a similar attribute where it can be cultivated. For Bajra it has 445 places where it can be cultivated. Again, for Urad it can be cultivated in 1220 places. For Groundnut, it can be cultivated in 620 places. Moreover, Rice can be cultivated in 2450 places. Arhar can be cultivated in 1200 places. Lastly Mung can be cultivated in 1610 places according to our collaborative filtering.

D. Fuzzy filtering

We had a total (3^3 = 27) combination because we have our 3 different input parameters: Rainfall, Humidity and Temperature and we divided them into 3 categories: Low (L),

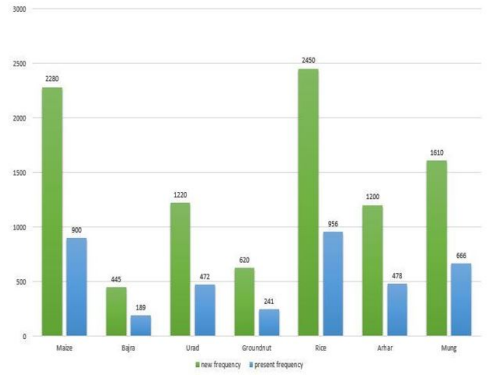

Fig. 11: Bar Chart of co-sine fitting of the different kinds of crops in Collaborative filtering

Moderate (M) and High (H). While filtering our dataset we got output on only 5 combinations:

Fuzzy combinations	Recommended crops
Rainfall = low Humidity = low Temperature = low	Maize, Rice, Wheat
Rainfall = low Humidity = low Temperature = moderate	Rice, Wheat, Gram
Rainfall = low Humidity = low Temperature = high	Maize, Rice, Arhar/Tur
Rainfall = low Humidity = moderate Temperature = moderate	Sunflower
Rainfall = moderate Humidity = low Temperature = moderate	Rice, Jute, Maize

TABLE II: Fuzzy Filtration

VIII. Conclusion

In this chapter future works and a brief review of current work has been discussed.

In future we have a hope that we design a web framework through Python Flask where it is easy to put input and see the desired result. We plan to use Python Flask because it can be accessed by the user and it helps the user to find the desired outcome and their expectations.

We also have a hope to increase our dataset, for now we have only worked on external factors such as rainfall, humidity, temperature. In future we want to include more factors in our dataset such as the fertility of the land. We want to predict possible diseases or the probable attacks of pests that can affect the production with sufficient parameters. We also add more crops to our recommendation as it will suggest the cultivators to go for much better alternatives.

To conclude, for our problems we have applied the algorithm for prediction and recommendation. For prediction we have applied the algorithm Naïve Bayes Classifier (48% for Rice), Support Vector Machine (69% for Rice), K-Nearest Neighbor (43% for Rice), Logistic Regression (37% for Rice), Random forest (36% for Rice). In this way we have predicted the productivity where the crop rice shows the maximum in every algorithm. And for recommendation using collaborative filtering it shows the recommended crops Rice, Maize, Moong as three most effective crops. On the other hand, fuzzy logic filters the crops according to the condition of the crop yielding factors. In this way we have worked on our ideas to develop. Our system solves the problems with limited parameters. So, increasing the parameters and adding more attributes to our system will be credible for the problems which were mentioned before. In this way we hope for the advancement of our agricultural economy, increasing productivity and contribution to the prosperity of the country.

Acknowledgment

First and foremost, we are grateful to God the almighty, to Whom the wisdom and power belongs, for His blessings and love to His creations.

Our appreciation goes to our thesis supervisor Dr. Muhammad Iqbal Hossain for granting a wonderful opportunity to work with him and to support and guide us in this thesis.

We are also indebted to everyone who helped us to proceed in our thesis work.

References

[1] D.S. Zingade, O. Buchade, N. Mehta, S. Ghodekar, and C. Mehta, "Crop prediction system using machine learning," *International Journal of Advance Engineering and Research Development Special Issue on Recent Trends in Data Engineering*, vol. 4, no. 5, pp 1-6, 2017.

[2] S. S. Dahikar and S. V. Rode, "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach," *International Journal Of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, vol. 2, no. 2, January 2014.

[3] V. Lamba and V. S. Dhaka, "Wheat Yield Prediction Using Artificial Neural Network and Crop Prediction Techniques (A Survey)," *International Journal For Research In Applied Science and Engineering Technology (IJRASET)*, vol. 2, no. 9, September 2014.

[4] N. Kim, K. J. Ha, N. W. Park, J. Cho, S. Hong, and Y. W. Lee, "A comparison between major artificial intelligence models for crop yield prediction: Case study of the midwestern united states, 2006–2015," *ISPRS International Journal of Geo-Information, publisher Multidisciplinary Digital Publishing Institute*, vol. 8, no. 5, pp 240, 2019.

[5] T. Sujjaviriyasup and K. Pitiruek, "Agricultural Product Forecasting Using Machine Learning Approach," *Int. Journal of Math. Analysis*, vol. 7, no. 38, pp 1869–1875, 2013.

[6] K. Moharm, "A Framework for Adaptive Personalized E-learning Recommender Systems," *International Journal of Intelligent Information Systems*, vol. 8, no. 1, pp 12-17, 2019.

[7] R. V. Meteren and M. V. Someren, "Using content-based filtering for recommendation", *Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 Workshop*, vol. 30, pp 47-56, 2000.

[8] A. A. Chirde and U. K. Biradar, "A Survey on Collaborative Filtering in Accordance with the Agricultural Application," *International Journal of Computer Applications*, vol. 975, pp 8887, 2014.

[9] M. Kuanr, B. K. Rath and S. N. Mohanty, "Crop Recommender System for the Farmers using Mamdani Fuzzy Inference Model," *International Journal of Engineering Technology*, vol. 7, 2018.

[10] P. Kaviani, S. Dhotre, "Short Survey on Naive Bayes Algorithm," *International Journal of Advance Engineering and Research Development*, vol. 4, no. 11, November 2017.

[11] N. Donges, *The Logistic Regression Algorithm*, January 08, 2019. Accessed on: Mar. 20, 2020. [Online]. Available: https://www.experfy.com/blog/the-logistic-regression-algorithm

[12] N. S. Chauhan, *Real world implementation of Logistic Regression*, March 11, 2019. Accessed on: Mar. 18, 2020. [Online]. Available: https://towardsdatascience.com/real-world-implementation-of-logistic-regression-5136cefb8125

[13] A. Navlani, *KNN Classification using Scikit-learn*, August 02, 2019. Accessed on: Mar. 13, 2020. [Online]. Available: https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn

[14] A. Navlani, *Support Vector Machines with Scikit-learn*, December 27, 2019. Accessed on: Mar. 13, 2020. [Online]. Available: https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python"

[15] M. R. Nikooa, R. Kerachian, M. R. Alizadeh, "A fuzzy KNN-based model for significant wave height prediction in large lakes," *Oceanologia*, vol. 60, no. 2, pp. 153-168, April 2018.

[16] Z. Cai, J. Gu, C. Wen, D. Zhao, C. Huang, H. Huang, C. Tong, J. Li, and Chen, H., "An intelligent Parkinson's disease diagnostic system based on a chaotic bacterial foraging optimization enhanced fuzzy KNN approach," *Computational and mathematical methods in medicine*, 2018.