Crop yield Prediction Using Machine Learning Algorithm.

Nikita S.Sapike
Department of Computer Engineering
Pimpri Chinchwad College of Engineering, Nigdi, Pune

Prof.S.S.Sambare
Department of Computer Engineering
Pimpri Chinchwad College of Engineering, Nigdi, Pune

***Abstract*—An important issue for the purposes of agricultural planning is a reliable yield estimate for the many crops involved in the planning. Machine learning is an approach to provide practical and efficient solutions to this problem. Many comparisons of ML methods for yield prediction have been made for the most accurate technique. Generally, the number of evaluated crops and techniques is too low and does not provide proper information for agricultural planning purposes. This paper compares the predictive accuracy of ML algorithm for crop yield prediction. People of india are practicing agriculture for years but the results are never satisfying due to various factors that affect the crop yield. To fulfill the needs of around 1.2 billion people, it is very important to have a good yield of crops. Due to factors like soil type, precipitation, region, seed quality, season, lack of technical facilities etc. The crop yield is directly influenced. Hence, new technologies are necessary for satisfying the growing need and farmers must work smartly by opting new technologies rather than going for trivial methods.In this paper, an Association Rule Mining technique integrating features of the Eclat algorithm and Genetic algorithm into the proposed method. The idea is to use the eclat technique of association rule mining to create rules and to use genetic algorithms to further refine those rules.**

***Index Terms*—Apriori algorithm, classification , Association Rule Mining technique, Machine Learning**

I. INTRODUCTION

Crop yield prediction is a major problem in agriculture. Starting each growing season, agricultural planners require estimating the yield for all the involved crops. Crop yield prediction is difficult because it depends on many interrelated factors. Moreover, yield is also affected by farmer decisions such as applied irrigation, pest and fertilizers, crop rotation, land type, and in controllable factors such as weather, season, subsidies and market. So, yield prediction traditionally has relied on farmers long-term experience for specific fields, crops and climate conditions, which can be inaccurate. Currently existing programs provide recommendation based on previously collected information from users. Such knowledge reflects the interests of a customer but at a given time it does not include intent. So, often existing systems don't produce the correct recommendation. General suggested systems work on collected information using explicit and implicit methods [1], as is the case with existing systems. Genetic recommendations offer recommendations in real time, using a fitness feature to estimate the suitability of recommended lists. Mining an association law. A technique combining the features of the Eclat and Genetic algorithms is used to evaluate the agricultural data set in order to produce recommendations provided by farmers on the basis of this data set for the future.The idea is to apply an Association Rule Mining technique, Eclat to generate rules and use Genetic Algorithms to further refine those rules and establish a relationship between them. Results show that the eclat-genetic algorithm model provides more accurate results compared to the traditional apriori algorithm.

II. REVIEW OF LITERATURE

In paper[1] discusses and compares the various data mining techniques for the decision support systems. The aim of this research is to implement a decision support system to predict the crop yield prediction from the collection of past data. In this paper, comparison of various data mining machine learning techniques are made for the smaller datasets and found that higher the accuracy. In paper [2] focuses on implementing crop yield prediction system by using data dining techniques by analysis on agriculture dataset. With the help of this information the percentage of loses and unsatisfactory yield will decrease as the management of the whole process and real statistics.

Paper[3] includes various data mining strategies for the use of fertilizer recommendations on agricultural soil datasets. Focused on different soil parameters such as the values Fe, S, Zn, Cu, N and Ph etc.This paper provides data mining techniques used in agriculture soil data and also compares the J48, Na¨ıve Bayes, JRip, K-Means classifier algorithm and suggest that an Artificial Neural Network is best as compare to other classification algorithms in data mining. Environmental [4] factors such as rainfall, temperature, soil type ,its chemical composition and total production taking into consideration and all the affecting parameters for the better selection of crop which can be grown over the season. In this paper successfully integrating machine learning with agriculture in predicting crop diseases, different irrigation patterns.

Indian farmers are [5] who do not choose the right crop based on their soil needs, since it affects productivity. Precision farming solved this problem. A soil database collected from the field, an agricultural crop, the attainment of parameters

such as soil through the soil testing laboratories dataset characterize this approach.Soil testing laboratory data provided by Using vector support machine and ANN as learners to recommend a crop for a given parameter to the recommendation system.

Paper [6] focuses on agricultural data analysis and identifying optimal parameters for optimizing crop production using data mining techniques such as PAM, CLARA and DBSCAN, Multiple Linear Regression to achieve the optimal climate requirement of wheat such as optimal temperature range, lowest temperature and falling rain for higher wheat crop production. DBSCAN gives the better clustering quality than PAM and CLARA. Precision agriculture [7] is a modern farming technique that uses soil characteristics, soil types, crop yield data collection research data and recommends the right crop to farmers based on their site specific parameters. This growing a crop's wrong choice, and increases productivity. In this paper, this problem is solved by recommendation system through an ensemble model with majority voting technique using Random tree, K-Nearest Neighbor and Naive Bayes. In [8] this paper has demonstrated the potential of various classification techniques and improve DSS for important prediction of crop yield productivity.

III. PERFORMANCE OF PROPOSED SYSTEM WITH EXISTING SYSTEMS

In existing system we does not have any computerizes system to recommend the crop to the farmer. [1] So farmer cannot understand which crop is take in which season . Due to rainfall or weather condition farmer is already in loss . If any natural disaster is came then whole farm is damage so his annual budget is collapse and farmer can suicide.Farmer does not understand which land is appropriate for which crop so he take huge production in his own land and increase the benefit.In proposed system providing solution to the farmer for the crop prediction. Take data base area wise ,season wise and other conditions and train them using machine learning algorithm and use data mining for data classification and machine learning technique for prediction And predict the crop area wise to the farmer.

IV. PROPOSED SYSTEM

Set Theory:

S = I, Fm, O, S, F I = {I1}.....set of Input.

I1 = Location of user

Fm = { GetLocation(), GetAttributes(latitude, longitude), GetSoil(), GetWeather(), FeasibleCrop(soil,weather), PastProduction(), ProfitableCrop(FeasibleCrops, PastProduction) MaxProfitableCrops() .}..... Set of functions.

Where, soil – N, P, K components

weather – Temperature and Rainfall values

O = Crop predicted for given Location . . . . Set of output.

S= Correct prediction for High production and profit

. . . . . Success

Condition

F = Failure in prediction due to incorrect training data

. . . .Failure Condition

This system takes into consideration the data related to soil, weather, region, season and past year production and suggests which are the best profitable crops which can be cultivated in the apropos environmental condition. As the system lists out all possible crops, it helps the farmer in decision making of which crop to cultivate. Also, this system takes into consideration the past production of data which will help the farmer get insight into the demand and the cost of various crops in market. As maximum types of crops will be covered under this system, farmer may get to know about the crop which may never have been cultivated.In the proposed system approri algorithm takes feedback from farmers like conditions of real-time crop area, season, weather conditions temperature etc. Approri algorithm operates on item that matches frequently. It count the minimum support count from the dataset.Then system calculate individual count from each region and then merge the condition like two forms and make group of itemset and calculate support of each group.Then calculate the percentage of each group and recommend the farmer to who has the highest percentage crop.

V. SYSTEM ARCHITECTURE

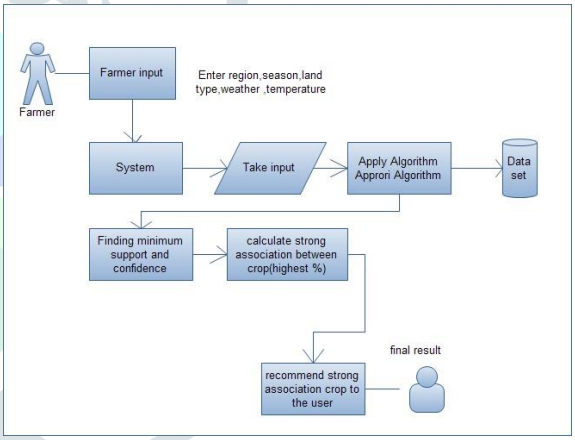

Fig. 1. System Architecture

The main goal to design proposed system is recommendation the crop according to surrounding condition for increase productivity of crop yield. In this system first farmer enter the conditions weather, land types, reason, season, for selected crop. By considering this parameter farmer can select proper crop in suitable season. In figure [1] show there are various conditions farmer can be select like weather, land, temperature, season, region using machine learning. In this system machine learning techniques are made for the smaller datasets and found that higher the accuracy higher will be rate for crop yield prediction. By making use of large datasets, one can improve the results. Machine learning is a method that helps discover patterns automatically from a large amount of data set in order to predict the proper outcome of unknown observations

based on previously identified patterns data set. By using this system we can predict the related crop to the farmer. In this system make data set region wise, season wise temperature and land data and used machine learning algorithm to recommend the crop for suitable for selected land and region. So farmer can increase income from the field. Farmer does not know about the surroundings condition like temperature,weather, and rainfall so using machine learning we are providing crop recommendation system for the prediction of crop is dependent on numerous factors such as Soil Nutrients, weather and past crop production in order to predict the crop accurately. All these factors are location reliant and thus the location of user is taken as an input to the system.

VI. FLOWCHART OF PROPOSED SYSTEM

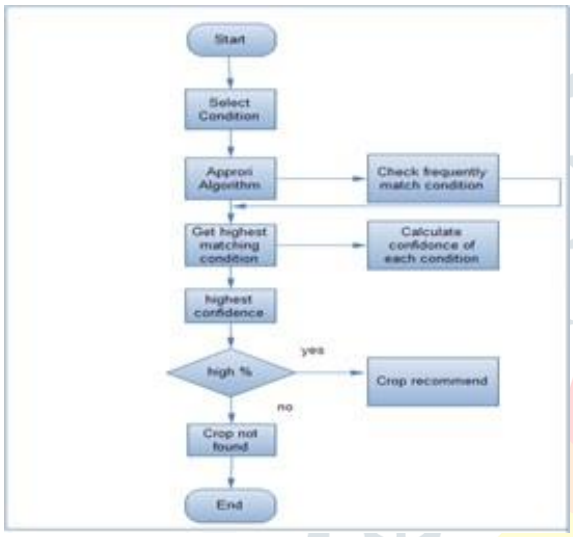

Fig. 2. Flow Diagram Of System

VII. ALGORITHM : APPRORI ALGORITHM

In this system an association rule mining algorithms are applied for the preprocessed data to get frequent itemsets and strong association rules are generated from frequently obtained itemsets. To only display those rules with the highest confidence value for each of the low, medium and high discrete values.

Apriori algorithm is used to find frequent patterns among items in the transactions stored in the database, and frequent itemsets are used to generate association law. This works mainly on property that ”All non-empty subsets of a common set of products must also be regular.Eclat algorithm is used to find frequent patterns among items in transactions stored in the database, and frequent itemsets are used to generate association law. It uses intersection-based approach for calculating the count of support for each candidate item. From the previous candidate itemset, frequent itemset is generated by pruning those items whose support does not meet minimum support threshold.

All non-empty subset of frequent itemset must be frequent. The key concept of apriori algorithm is its anti-monotonicity of support measure. Consider the following example given from dataset to find frequent itemsets and generate association rules for them.

Step:1

Conditions	Frequently Match Item
(I1)Pune	4
(I2)Summer	6
(I3)irrigated	3
(I4)Cloudy	7
(I5)20-30	9

TABLE I
ITEM TABLE

Step :2

Conditions	Frequently Match Item
(I1)(I2)	2
(I2)(I3)	3
(I3)(I1)(I2)	1

TABLE II
ITEM TABLE

Step :3

Conditions	Frequently Match Item
(I1)(I2)(I3)	2
(I2)(I3)(I4)	3

TABLE III
ITEM TABLE

Minimum support count =3
minimum confidence is 60%

A confidence of 60% means that 60% of the condition having the same crop recommendation ..
Confidence(A>B)=Support_count(A? B)/Support_count(A)
So here, by taking an example of any frequent itemset, we will show the rule generation.
Itemset {I1, I2, I3} //from L3
SO rules can be
[I1^I2]=>[I3] //confidence = sup(I1^I2^I3)/sup(I1^I2) = 2/4*100=50%
[I1^I3]=>[I2] //confidence = sup(I1^I2^I3)/sup(I1^I3) = 2/4*100=50%
[I2^I3]=>[I1] //confidence = sup(I1^I2^I3)/sup(I2^I3) = 2/4*100=50%
[I1]=>[I2^I3] //confidence = sup(I1^I2^I3)/sup(I1) = 2/6*100=33%
[I2]=>[I1^I3] //confidence = sup(I1^I2^I3)/sup(I2) = 2/7*100=28%
[I3]=>[I1^I2] //confidence = sup(I1^I2^I3)/sup(I3) = 2/6*100=33%

So if minimum confidence is 50%, then first 3 rules can be considered as strong association rules.

Means our condition match the three rows of dataset in table. Then
system recommend the most highest percent accuracy crop like
sugarcane, cotton, bajara .

VIII. DATASET

For this System used real time data set to calculate accuracy for crop yield prediction.For dataset collection, Consider the region list in Maharashtra like Pune, Solapur, Nashik,Ahemadnagar, Consider seasons like summer, winter, rainy also Collect land types in maharashtra.Then consider temperature and weather conditions like sunny ,cloudy, forecast. Then take conditions like region, season, weather condition, temperature, land types in real time. Which can calculate the appropriate algorithm-based crop-prediction model. In this system apriory algorithm finds the element that matches itemset in the dataset and finds the least support between them. And then find confidence in each situation for finds a suitable model of crop prediction.Following figure show the example of dataset.

Id No	Region	Season	Land Types	Weather	Temprature	Crop type
1	Pune	Rainy	Bagayat or irrigated land	Cloudy	20-30	Cotton
2	Pune	Rainy	Bagayat or irrigated land	Sunny	20-30	SugerCane
3	Pune	Rainy	Bagayat or irrigated land	Overcast	20-30	SugerCane
4	Nashik	Rainy	Bagayat or irrigated land	Cloudy	20-30	SugerCane
5	Nashik	Rainy	Bagayat or irrigated land	Sunny	30-40	SugerCane
6	Nashik	Rainy	Bagayat or irrigated land	Cloudy	30-40	SugerCane
7	Nashik	Rainy	Bagayat or irrigated land	Sunny	20-30	Cotton
8	Nashik	Rainy	Bagayat or irrigated land	Sunny	30-40	SugerCane
9	Nashik	Rainy	Bagayat or irrigated land	Overcast	30-40	Cotton
10	Nashik	Rainy	Bagayat or irrigated land	Sunny	20-30	SugerCane
11	Nashik	Rainy	Bagayat or irrigated land	Cloudy	30-40	Cotton
12	Nashik	Winter	Bagayat or irrigated land	Sunny	30-40	SugerCane
13	Nashik	Winter	Bagayat or irrigated land	Overcast	20-30	Cotton
14	Nashik	Winter	Bagayat or irrigated land	Cloudy	30-40	SugerCane
15	Nashik	Winter	Bagayat or irrigated land	Overcast	20-30	Cotton
16	Nashik	Winter	Bagayat or irrigated land	Sunny	30-40	SugerCane
17	Nashik	Winter	Bagayat or irrigated land	Overcast	20-30	Cotton

Fig. 3. Sample Data Set

IX. RESULT AND DISCUSSION

Association rules find all sets of items (itemsets) that have support greater than the minimum support and then using the large itemsets to generate the desired rules that have confidence greater than the minimum confidence. The lift of a rule is the ratio of the observed support to that expected if X and Y were independent. Association rules are generally used for clustering, in crop dataset it is used for the clustering of crops in the respective clusters.Association rules generated for the crop dataset using WEKA tool is shown in the figure given below.

Fig. 4. Association Rule

Approri Algorithm Accuracy Accuracy is one metric for evaluating classification models.
Informally, accuracy is the fraction of predictions our model got right. Formally, accuracy has the following definition:

Accuracy=Number of correct predictions/Total number of predictions.

For binary classification, accuracy can also be calculated in terms
of positives and negatives as follows:
Accuracy=TP+TN/TP+TN+FP+FN.

Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

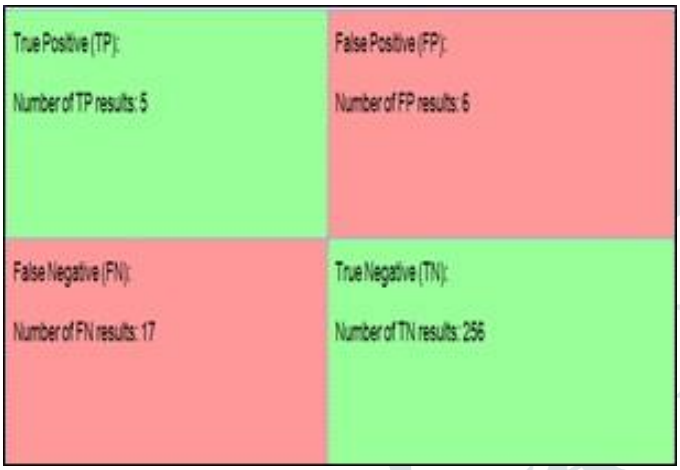

Fig. 5. Confusion Matrix

Accuracy=TP+TN/TP+TN+FP+FN=5+256/5+256+6+17.
Accuracy=261/284=0.9190.
Accuracy comes out to 0.91, or 91%. That means our prediction algorithm is doing a great job of identifying crop prediction.

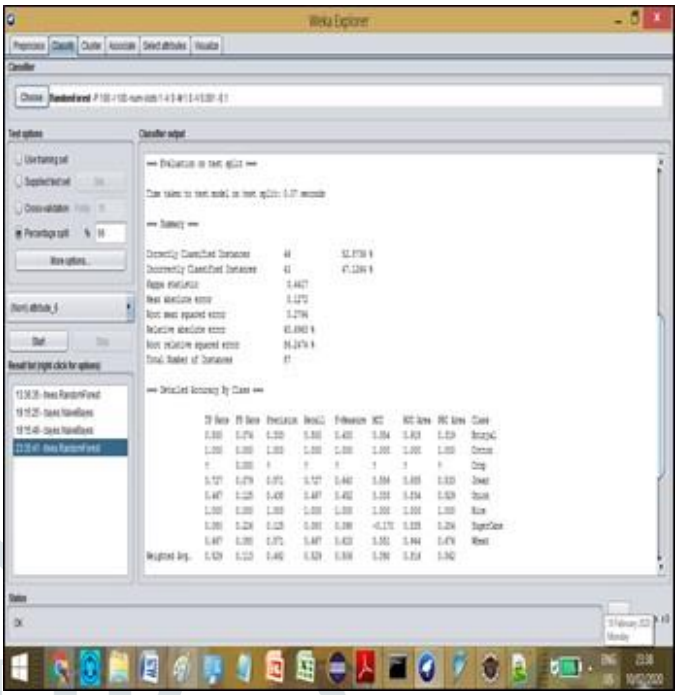

Fig. 6. Comparative Analysis

A. Comparative Analysis

Random Forest is a bagging technique which is based on tree ensemble machine learning method. It generates multiple tree of randomly sub-sampled features. The output of forest is evaluated by taking average value of the prediction of individual trees. Since it is using random sub-sampled features, random forest can be used in high dimension input predictor[7]. Following are the results of random forest algorithm for crop prediction when applied on crop dataset on WEKA tool.

Naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. Depending on the precise nature of the probability model, Naive Bayes classifiers can be trained very efficiently in a supervised learning setting. The problem of predicting the crop is formalized as a classification rule, where Naive Bayes is used[9]. The results generated for crop dataset by applying Na¨ıve Bayes classifier using WEKA tool is shown in the given figure.

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute	Class: Brinjal (0.12)	Cotton (0.16)	Crop (0)	Jowar (0.12)	Onion (0.16)	Rice (0.07)	SugarCane (0.26)	Wheat (0.12)
attribute_1								
Ahemednagar	7.0	21.0	1.0	7.0	9.0	1.0	1.0	7.0
Nashik	7.0	1.0	1.0	7.0	1.0	1.0	19.0	7.0
Pune	7.0	1.0	1.0	7.0	33.0	19.0	33.0	7.0
Region	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Sangali	10.0	1.0	1.0	7.0	1.0	1.0	19.0	7.0
Solapur	4.0	21.0	1.0	7.0	1.0	1.0	1.0	7.0
[total]	36.0	46.0	6.0	36.0	46.0	24.0	74.0	36.0
attribute_2								
Rainy	31.0	5.0	1.0	1.0	11.0	7.0	21.0	1.0
Summer	1.0	25.0	1.0	31.0	13.0	7.0	25.0	1.0
Weather1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Winter	1.0	13.0	1.0	1.0	19.0	7.0	25.0	31.0
[total]	34.0	44.0	4.0	34.0	44.0	22.0	72.0	34.0
attribute_3								
Bagayat	31.0	1.0	1.0	31.0	19.0	1.0	55.0	31.0
Jirayat	1.0	1.0	1.0	1.0	23.0	1.0	15.0	1.0
Land	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Riceland	1.0	1.0	1.0	1.0	1.0	19.0	1.0	1.0
Warkas	1.0	41.0	1.0	1.0	1.0	1.0	1.0	1.0
[total]	35.0	45.0	5.0	35.0	45.0	23.0	73.0	35.0
attribute_4								
Cloudy	16.0	7.0	1.0	1.0	13.0	7.0	22.0	16.0
Overcast	1.0	19.0	1.0	16.0	16.0	7.0	25.0	16.0
Sunny	16.0	17.0	1.0	16.0	14.0	7.0	24.0	1.0
Weather2	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
[total]	34.0	44.0	4.0	34.0	44.0	22.0	72.0	34.0

Fig. 7. Navie Bayes Classifier

After analysing the results of Na¨ıve Bayes, Random Forest, Gradient Boosted Trees and Support Vector Machine and Apriori algorithm for crop prediction using the RapidMinner and Weka Tools, it can be said that crop prediction can be efficiently done using Apriori Algorithm which gives the result as follows:

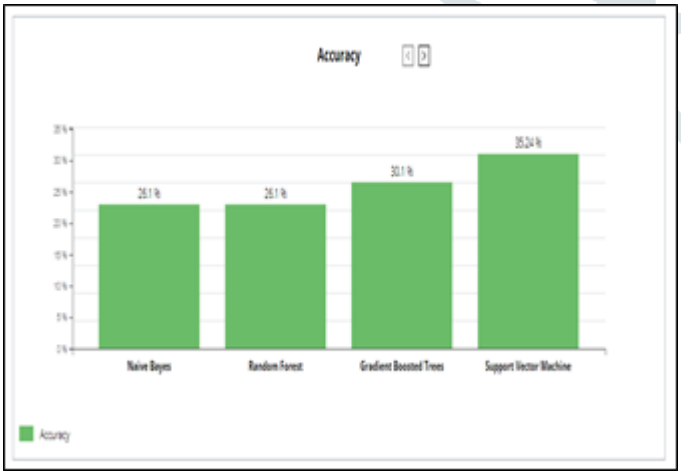

Fig. 8. Accuracy Graph representation

ALGORITHMS	RESULTS
NAÏVE BAYES	26.1%
RANDOM FOREST	26.1%
GRADIENT BOOSTED TREES	30.1%
SUPPORT VECTOR MACHINE	35.24%
APRIORI	91.90%

Fig. 9. Accuracy table represenatation

Approri algorithm are used in proposed system for crop prediction. Approri algorithm tests all the element in the dataset that are always matches. In the dataset it checks all element one by one and Join the parameter in two ways afterwards and find frequency.In last step we count all frequency of all input data parameters and estimate the specific performance of the highest matching number of parameters in input data. Approri estimate better output than other algorithms such as linear regression, multiregression, and other algorithms for data mining. Based on string dataset approri algorithms that's more reliable than other algorithms.

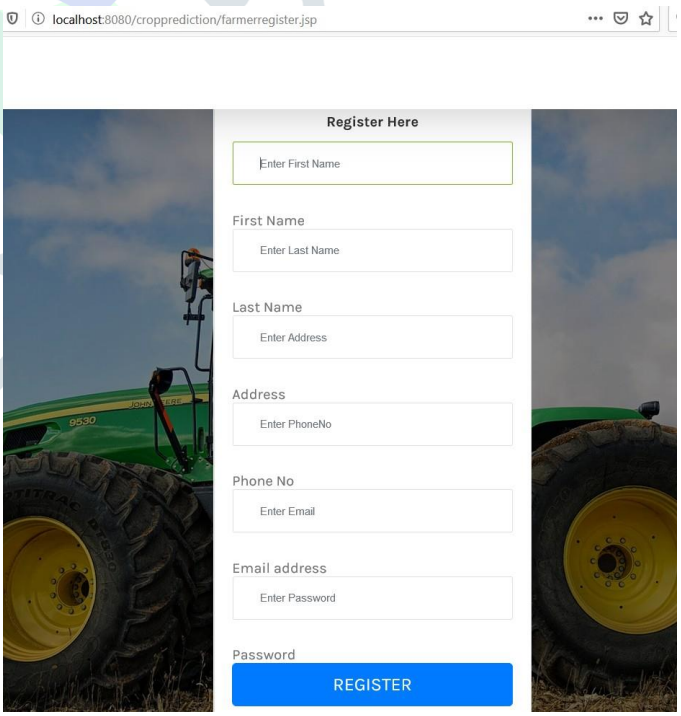

Fig. 10. Farmer Registration Form

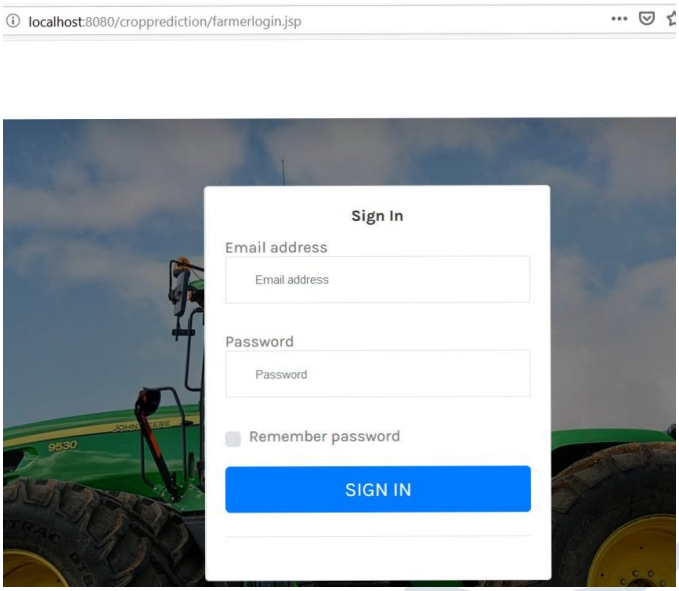

Fig. 11. Login Form

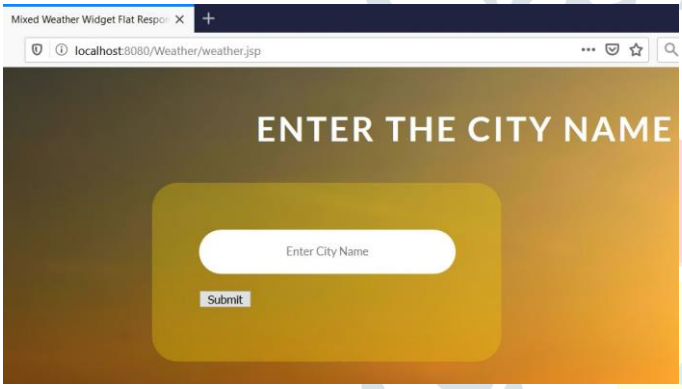

Fig. 12. City Name For Checking Weather Condition

Fig. 13. Weather Condition

X. Conclusion

We observe that, after analyzing the results of Na¨ıve Bayes, Random Forest, Gradient Boosted Trees, Support vector machine and Apriori algorithm for crop prediction using Weka Tools, it can be said that crop prediction can be efficiently done using Apriori algorithm which give the more accuracy than other algorithm.This system achieve an accuracy 91%. The system work takes into consideration the related data to soil, weather, season, region, and suggests which are the best profitable crops which can be cultivated in environmental condition. This system will help farmers to increase productivity in agriculture.

XI. Future Scope

In the future, all farming devices can be connected over the internet using IOT. The sensors can be employed in farm which will collect the information about the current farm conditions and devices can increase the moisture, acidity, etc.In the future, we can use real-time weather and soil data sets that are personally collected by equipment or the data sets. We can merge distinct classifiers to construct a single model called ensemble.

References

[1] Kusum Lata, Bhushan Chaudhari , “Crop Yield Prediction Using Data Mining Techniques And Machine Learning Models For Decision Support System” , JETIR April 2019.
[2] Shruti Mishra Priyanka Paygude,” Use of Data Mining in Crop Yield Prediction” ,ICISC 2018.
[3] Jignasha M. Jethva, Nikhil Gondaliya , Vinita Shah ,” A Review on Data Mining Techniques for Fertilizer Recommendation” , IJSRCSEIT 2018.
[4] Nishit Jain, Amit Kumar, Sahil Garud, Vishal Pradhan, Prajakta Kulkarni , “Crop Selection Method Based on Various Environmental Factors Using Machine Learning”, IRJET 2017.
[5] Rohit Kumar Rajak , Ankit Pawar, Mitalee Pendke, Avinash Devare, “Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique”, IRJET 2017.
[6] Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki , “Analysis of agriculture data using data mining techniques: application of big data” , Journal Big Data 2017.
[7] S.Pudumalar , E.Ramanujam , R.Harine, Rajashreeń, C.Kavyań, T.Kiruthikań, J.Nishań , “Crop Recommendation System for Precision Agriculture”, ICoAC 2016
[8] Ami Mistry, Vinita Shah , “Brief Survey of data mining Techniques Applied to applications of Agriculture ,international Journal of Computer Applications” IJARCCE 2016.
[9] Nikita Gandhi, Leisa J. Armstrong , “Rice Crop Yield Forecasting of Tropical Wet and Dry Climatic Zone of India Using Data Mining Techniques “, ICACA 2016.
[10] Bhuvana, Dr.C.Yamini (2015), “Survey on Classification Algorithms in Data mining”. International Conference on Recent Advances in Engineering Science and Management
[11] Monali Paul, Santhosh K. Vishwakarma, Ashok Verma, “Prediction of crop yield using Data Mining Approach” Computational Intelligence and Communication Networks (CICN), International Conference 12-14 Dec. 2015.
[12] Yogesh Gandge, Sandhya, “A Study on Various Data Mining Techniques for Crop Yield Prediction”, 2017 ICEECCOT.
[13] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh (2015), ’Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique’, ICSTM.
[14] Raorane A.A., Kulkarni R.V., ”Data Mining: An effective tool for yield estimation in the agricultural sector”. IJETTCS July August 2012.