TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 5776-5788

Multiple Crop Pest Classification Using Big Data Analytics

R.P.L.Durgabai[1],P.Bhargavi[2] S.Jyothi[3]

Department of Computer Science, Sri PadmavatiMahilaVisvavidyalayam, Tirupati
[1]poonamramchandra@gmail.com, [2]pbhargavi18@yahoo.co.in[2], [3]jyothi.spmvv@gmail.com[3]

Article Info
Volume 83
Page Number: 5776 - 5788
Publication Issue:
May - June 2020

Article History
Article Received: *19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *17 May 2020*

Abstract

Agriculture is a mixture of so many crops, like food and non-food. Food staples being rice and wheat. Indian farmers also grow pulses, potatoes, Sugar Cane. And non-food items are cotton, tea, coffee, rubber and jute. In Andhra Pradesh state Agriculture is an age old economic activity. Where many major crops are cultivated in this state, the largest problem faces in agriculture is pests attack. Pest attack not only causes loss for a crop but also affects the farmers to live a pathetic life. To classify the crops and pests a non-persistent based method like big data analytics is reliable and to analyse the data by using machine learning algorithms, which can give the efficient results. In this context, six popular machine learning algorithms like Decision Tree (DT), Random Forest Classifier (RFC), K-Nearest Neighbour (K-NN), Gaussian Naïve Bays (GNB), Support Vector Machine (SVM) and Ada Boost Classifier (ABC) are applied. Among them Random Forest Classifier(RFC) performed in time and given accurate values while compared with other methods. The Random Forest Classifier makes the agriculture sector to classify and comprehend the pests easily. The ML based pest classification system is helpful to the department of agriculture to analyse the pests in various crops and in various seasons.

Keywords: *Random Forest Classifier, Big Data, Machine Learning, crops, pests, classification, MongoDB.*

I. INTRODUCTION

Agriculture proves to be a major factor of Indian economy and it involves production of crops. Crops may be either food crops or commercial crops. Food crops containmillets, wheat, maize, paddy, grams, etc., howevermarketable crops are groundnut, cashew, cotton, sugarcane, and etc. The crop productivity is being influenced by weather conditions [1]. Hence, accurate yield prediction is a major problem that ought to be addressed. Early prediction of yield would facilitate the farmers to make precautionary actions to improve productivity. The recurrence of droughts and severe cyclonic storms are the major problems encountered. Early prediction is possible through collection of previous experience of the farmers, weather conditions and other influencing factors and; store it in a large database [2].

Every Year globally many livestock, fish harvests, crops are damaged due to pests and diseases changed in climate patterns results a new or early encountered disease and pest attack at unprepared regions with detrimental consequences to industries and livelihoods. The recurrence of droughts and severe cyclonic storms are the major problems encountered here. Suggested obstacles for natural resource management in A.P agriculture are water harvesting, site-specific nutrient management, amelioration of millions of hectares of land and protection of crops from pest attack in all seasons and in various temperatures [5]. The early warning is possible for pest and a disease outbreak is critical at given risk. Ground truth diseases are costly and time consuming. Therefore new technologies and approaches are very essential to help and monitor

May – June 2020
ISSN: 0193-4120 Page No. 5776-5788

the spread of out breaks and alert which are being affected.

Bigdata in agriculture is necessity for large investment to create infrastructure for data storage and processing [3]. This needs to operate the real time applications like monitoring the crops, pests, weather forecasting and animal disease. Hence big data analysis is used to describe a new generation of practices designed to extract economic value or organizations and farmers related. This has very large volumes and wide variety of data to enable the high- velocity capture, discover the analysis. Big data analysis is used in various industries like banking, online users, insurance and personalized in environment studies, government organizations etc., to enhance their ability to serve citizens addressing national challenges related to economy, job creation, health care.

The data collection, analysis and integration are easily calculated by an automated tool called Machine Learning. The day is saved by ML algorithm because if formed by combining cloud computing and machine learning. This will analyse the existing data and generate required classification [4].

II. LITERATURE SURVEY

This context describes about the literature survey of prediction of pest which are implemented by other research as follows: Dr. S. Hari Ganesh et.al. [6] The author given the overview of data mining techniques in agriculture. The crop prediction is done for production based on functions weather and soil conditions. These techniques used in agriculture field to increase the income for farmers to predict the climate data set.

U.Latha et.al [7] use the agricultural data to analyse and predict the accuracy of an algorithm. Here used Naive Bayes and J48 algorithm and got Navie Bayes is performed better than the decision tree algorithm J48.

Hetal Patel et.al [8] equated efficiency of the classification algorithms named J48, Simple Cart and Naive Bayes. By generated results the best algorithm for prediction is decision tree is J48 by 89.33%, and remaining got 85.66% and 82.66% respectively for the given data set.

Revathy R, et.al [9] proposed C4.5 algorithm for building decision tree to provide efficiency of C4.5 classifier which handle the crop pest data with missing values and eliminates over fitting while constructing the tree. But noted that the size of the data set is growing enormously big to overcome big data is used.

UmairAyub et.al [10] analysed different data mining classifiers with different feature sets to predict the grass grub damages. The author used the DT, RF, NN, GNB, SVM and KNN. RF and GNB performed other classifiers and also designed ensemble models by combining different classifiers to improve accuracy. By combining DT, RF AND SVM has proven best combination and achieving good results.

CH Chandrasekhar et al. [11] might scaled up in terms size of data and crop variations by that Big Data Analytics is the best system for crop planning to improve agriculture productiveness and very significant for agriculture manufacturing to offer income for ICT & too common man.

B.M. Sagar et.al [12] surveyed on new concepts like smart farming, digital agriculture, precision agriculture etc. and observed analysis on agriculture soils, crop yields, climatic conditions of the given data set. They also activates the numerous fields such as soil quality, crop yield, seed selection, weather forecasting etc. also identified the major trends and crop yield prediction.

Liakos Konstantinos et al [13] explained about applications of machine learning in agricultural, and production systems are discussed and explained about benefits of machine learning technologies in agriculture.

May – June 2020
ISSN: 0193-4120 Page No. 5776-5788

Ansif Arooj et al [14] discussed different data mining classification algorithms and applied this methods to data sets for classification of agricultural regions on the basis of soil properties.

Avinash Kumar et.al [15] evaluated the best suitable crop and which pest effects the crop at particular time to farmers and also suggests the pest control techniques to prevent the damage.

III. PROBLEM DOMAIN

A problem domain depends purely on the interest and capability of a person and application area which needs to have deeper study. The present problem describes the classification of pest in multiple crops all through different seasons and temperature.

Description of crop pest data

The dataset is collected from AP agriculture which is based on temperature and seasons. The data set has twenty major crops which are cultivated in A.P by considering all the four features like pest name, crop name, seasonal occurrence and temperature. Size of the data set has more than one million instances and does not have a single non available value. The target output label was extracted and used for classifying the pest name. The complete information of the dataset is as shown in below Table 1.

Table 1. A.P. Crops Pests Dataset

S.No.	Features
1	Pest name
2	Crop Name
3	Seasonal Occurrence
4	Temperature

The analysis is apprehended by taking machine learning approach for crop pest classification, which is applied to dataset. The classification task with pest analysis problem is assigned to a label of particular instance. High dimensional datasets have the problem of irrelevant or redundant features which often lowers the performance of machine learning algorithms [17] and the present data sets requires all the four features to make classification.

IV. METHODOLOGY

Methodology analyses the information which belongs to a particular area.

The proposed research is done based on multiple crop pests classification using cultivated crops in A.P. the system has trained and tested on A.P crop pest dataset. All the computations were performed in python.

The major contribution of the present work is illustrated in figure 1. The A.P crops pest dataset is selected and pre-processed. For the pre-processed dataset machine learning algorithms is applied to check the performance of all the attributes in terms of classification [16] accuracy and execution time. Based on the result analysis among all the classifiers RFC is considered as a best classifier for analysis as it helps for interpretation and visualization.

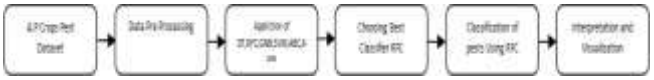

Fig 1. Proposed Classifier

MongoDB

The data set is validated to extract information from the dataset by constructing positive networks among the existing data, which enhances the scope of information. The amount of data that is visualized and analysed using big data is too voluminous for an analyst. So for that purpose MongoDB is used for sorting this enormous data which is highly flexible.

MongoDB is a database storage which can give high performance, constant availability and easy expandability. It is also known as NoSQL database because it is a cross platform document- oriented database system [17]. It provides a solution for different key- values and traditional RDBMS systems. MongoDB gains its popularity from the distributed key value store. Hence, MongoDB is services as the database for Big Data Processing [18].

Classification of crops pest

Classification can be performed on both structured and unstructured data. The classes can be categorized based on the procedures that we follow on the given data. The instance labels are categorized and recognized normally to describe attributes in a dataset. The dataset is grouped under multi-class classification. Machine learning algorithm is applied to classify the crop pest in different temperature and seasonal occurrences. By applying machine learning and data analytics combinly to make organizations grow by lessening the causes of potential risks, eliminates the failures in policies, purchase based time to time offers and striking out the fraudulent activities by cross checking the data. The classification fields in ML is either binary, multi-class, multi-labeled or hierarchical [19].

Machine learning

Data collection, analysis and integration can be done completely using ML algorithms. If the input information is very small and can be done manually there is no need of using ML. If the data adds on every day, and increases the data beyond the limit ML algorithm can be used for quick processing and through analysis. Data analytics can be used for any type of data such as descriptive, diagnostic, predictive, planning and classification [20]. Some popular classifications algorithms are applied on this dataset are discussed briefly.

Decision Tree

Where to build a tree Standard Deviation(SD) It is a form of flowchart structure in which internodes represents a feature and each leaf node represents a decision taken after computing all features. While some of the associated decision tree is incrementally developed [21].

$$Avg = \bar{x} = \frac{\sum x}{n} Count = n\ SD = S$$

$$= \sqrt{\frac{\sum(X - \bar{X})^2}{n}}\ CV = \frac{S}{\bar{X}} * 100\%$$

is used, to decide when to stop the branch Coefficient of Deviation (CV) is used, to calculate leaf node Average (Avg) is used and Count n is also used to derive the decision tree.

For two attributes target and predictor is evaluated by

$$S(T, X) = \sum_{c \in X} P(c) S(c)$$

Support Vector Machine

SVM is a branch of supervised machine learning algorithm which can be applied for both classification and prediction problems [22]. It producesan optimal hyper plane outcome to classify the datasets that supports the hyper plane is called support vector.

$$f(x) = W^T X + b$$

W is a normal line which is known as Weight vector and b as bias.

Gaussian Naive Bayes

Naive bayes is a classification algorithm applied for both binary and multi-class classification problems. It can be extended to real valued attributes, most generally by assuming a Gaussian distribution which is called as Gaussian Naïve Bayes. This is one of the best and easiest functions to estimate the distribution of the data [23].

May – June 2020
ISSN: 0193-4120 Page No. 5776-5788

$$\hat{p}\left(\frac{y}{\vec{x}}\right) = \frac{p(y \cap \vec{x})}{p(x)}$$

This can be estimated using the MLE method assuming y is discrete.

$$\hat{p}\left(\frac{y}{\vec{x}}\right) = \frac{\sum_{i=1}^{m} x(\vec{y}^{(i)} = y \cap \vec{x}^{(i)} = \vec{x}}{\sum_{i=1}^{m} x(\vec{x}^{(i)} = \vec{x})}$$

This algorithm estimates only good if there are many training vectors with the same identical features

K- Nearest Neighbour

The KNN algorithm is a supervised machine learning algorithm that can be applied in classification as well as regression problems. The main purpose behind nearst neighbor classification consists of finding a predefined number ie., K training samples closest in distance to a new sample [24]. The label of the new sample will be defined from these neighbours which have to be determined. Non generalizing machine learning are considered as neighbours based methods as they remember all of its training data [25].

Distance Function

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$ Manhattan $\sum_{i=1}^{k}|x_i -$

Random forest classifier

Random forest is a classifier composed of multiple decision trees. The category of each data object in the test set is determined by the mode of the category exported by each decision tree. The decision tree has tree structure where the leaf node is the final classification category and other nodes are attribute nodes used to determine the type of the data object to test data set [26]. When data object is given as input then it firstly determines the direction by attributes of the root node until it reaches the leaf node. The leaf node category of the data object set is tested and determined by the decision tree [27].

$$f(x) = \sum_{m=1}^{M} \beta_m h_m(X)$$

Where X is a predictor, m is the transformation of X, β_m is the weight given to the mthtransformation, h_m is the mth transformation of X, f (x) is the linear combination of transformed values of X.

Ada Boost Classifier

Ada Boost classifier combines weak classifier algorithm to make strong classifier. Sometimes a single algorithm may classify the objects poorlyand such individual classifiers are combined with multiple classifiers with selection of training set at every iterations and right amount of weight is assigned in final voting. It will give good accuracy score for overall classifier [28,29,30].

$$F(x) = sign\left(\sum_{m=1}^{M} \theta_m f_m(x)\right)$$

Where f_m stands for mth weak classifier, θ_m is corresponding weight

V. EXPERIMENTAL RESULTS

Initially, install all the required packages in python software because it is a general purpose software with high level programming language due to it has wide selection of libraries framework to improve leverage theme.

The dataset is loaded in MongoDB for storage of large data and it is retrieved in python program. The current dataset is very large besides having multi class labels to make clear analysis and to generate rules for interpretation and visualization Machine learning algorithms are applied to classify the crop pest dataset on different pest occurrences based on climatic conditions. Each classifier accuracy,

May – June 2020
ISSN: 0193-4120 Page No. 5776-5788

percentage and time taken to predict each model is compared as shown in table 2. By comparing K-NN,GNB,DT,SVM,RFC,ABC K-NN algorithm is optimal in giving good accuracy and less time to predict, it is not considered as suitable for present dataset because it unable to visualize and interpret.

So, RFC algorithm is used to build single classification tree very easily. Random Forest include too many classification trees. To group a particular object under any of the tree we do classification. Considered N cases as sample, replace the original data will act as the training set for the tree which is growing. If the number of input variables is more and is less than the specified nodes then the best split can be made exactly at M. Further this value has maintained constantly throughout the growth of the tree. Pruning has to be avoided even through the tree is growing to the maximum extent. Using the total dataset ten trees have been generated. Seasonal occurrence and temperature are considered as the basic criteria for generating the tree. Depending upon the results drawn in the tree we can conclude which pests is more prone for attack on the crop at which seasonal occurrence and temperature.

Table 2. Comparative study of Classification algorithms

S. No	Classifier Applied	Accuracy Percentage	Model Prediction Time in seconds	Rules generated	Interpretation	Visualization
1	K-NN	95	0.24	No	No	No
2	GNB	70	0.9	No	No	No
3	DT	38	60	No	No	No
4	SVM	75	0.01	No	No	No
5	RFC	95	90	Yes	Yes	Yes
6	ABC	80	78	No	No	No

Test results and analysis

X0 is Seasonal Occurrence, X1 is Temperature

Tree 0

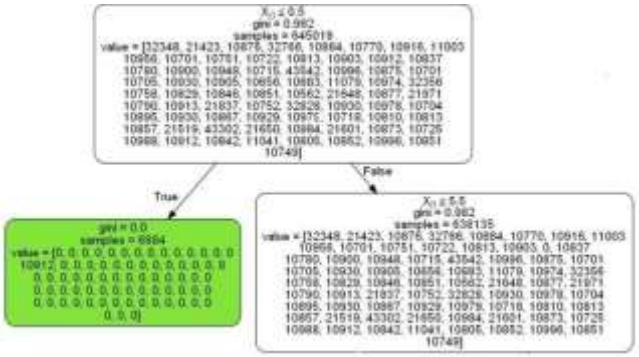

Fig 2. Tree based on Seasonal Occurrence

Rule 1: If it is a dry season then maruc pod borer pest attacks high on black gram crop.

Rule 2: If rule 1 is true then spodopteraexigua pests occurs high.

Rule 3: If rule 1 is false then during spring season maruc pod borer attacks high on black gram crop.

Rule 4: If it is rainy season then green leafhopper pest attacks high on rice crop.

Rule 5: If temperature is less than equal to 30 degrees then maruc pod borer pest attacks high on black gram crop.

Rule 6: If temperature is less than 23 degrees then stem weevil pests attacks high on jute crop.

Rule 7: If there is warm season then mites pests attacks high on cotton, jowar, jute, sesame and chilly crops.

Rule 8: If temperature is less than equal to 31 degrees then climbing cutworm attacks high on rice crop.

Rule 9: If temperature is less than equal to 27degrees then seedling blight pest attacks high on jute crop.

Rule 10: If temperature is less than equal to 35 degrees then leaf hopper pest attacks cotton crop.

May – June 2020
ISSN: 0193-4120 Page No. 5776-5788

Rule 11: If temperature is less than equal to 24 degrees then mites' pest attacks on cotton, jowar, jute, sesame and chilly crops.

Tree 1

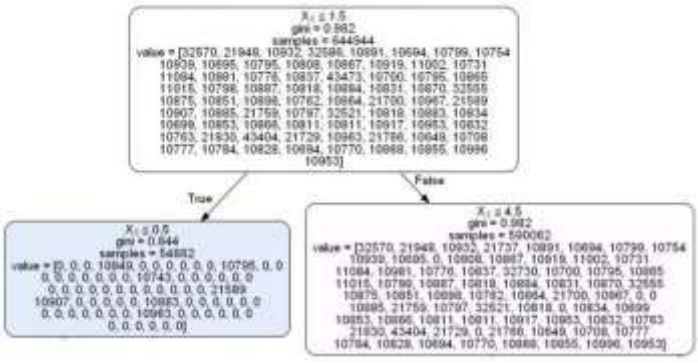

Fig 3. Tree based on Temperature

Rule 1: If temperature is less than equal to 21 degrees then maruc pod borer pest attacks on black gram crop.

Rule 2: If rule 1 is true and temperature is less than equal to 20 degrees then mite pest attack high on jowar crop.

Rule 3: If rule 1 is false and temperature is less than equal to 24 degrees then ear head caterpillars' pests attack high on ragi crop.

Rule 4: If temperature is less than equal to 23 degrees then mites pest attack on cotton, sesame and chilly crops.

Rule 5: If temperature is less than equal to 22 degrees then ear head bug pest attacks on jowar crops.

Rule 6: If it is spring season then leaf webber cum capsule borer pest attacks on sesame crop.

Rule 7:If it is spring season then leaf folder pest attacks on soybean and rice crops.

Rule 8: If it is spring season then ear head caterpillars' pest attacks on ragi crop.

Rule 9: If temperature is less than equal to 27 degrees then maruc pod borer pest attacks on black gram crop.

Rule 10: If temperature is less than equal to 25 degrees then stem weevil pest attacks on Jute crop.

Rule 11: If temperature is less than equal to 36 degrees then stemborers pest attacks on rice, jowar and ragi crops.

Rule 12: If it is warm season then stemborers pest attacks on rice, jowar and ragi crops.

Tree 2

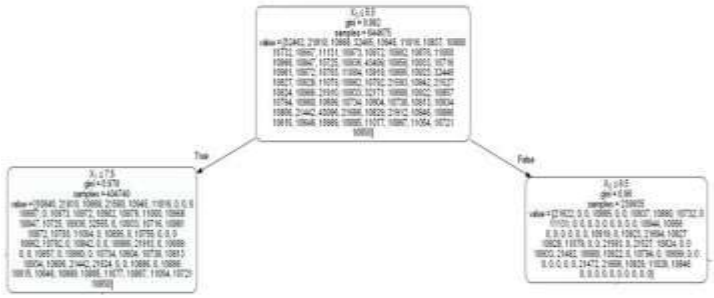

Fig 4. Tree based on Seasonal Occurrence

Rule 1: If it is spring season then maruc pod borer pest is attacking high on black gram crop.

Rule 2: If rule 1 is true and if temperature is less than 27 degrees then maruc pod borer pest is attacking on black gram crop.

Rule 3: If rule 2 is false and if it is warm season then mites pest attacks chilly, cotton and sesame crops.

Rule 4: If temperature is less than equal to 36 degrees then maruc pod borer pest is attacking high on black gram crop.

Rule 5: If temperature is less than equal to 28 degrees then ear head caterpillars' pest is attacking high on ragi crop.

May – June 2020
ISSN: 0193-4120 Page No. 5776-5788

Rule 6: If temperature is less than equal to 35 degrees then cut worms' pest attack korra crop.

Rule 7: If temperature is less than equal to 23 degrees then leaf folder pest attack high on soybean and rice crop.

Rule 8: If temperature is less than equal to 26 degrees then stem weevil pest attack high on jute crop.

Rule 9: If temperature is less than equal to 24 degrees then aphids pest attack high on black gram, cotton, chilly.

Rule 10: If temperature is less than equal to 31 degrees then seedling blight pest attack high on jute crop.

Rule 11: If temperature is less than equal to 35 degrees then leaf hopper pest attack high on cotton crop.

Rule 12: If temperature is less than equal to 27 degrees then pink bollworm pests attack high on cotton crop.

Rule 13: If temperature is less than equal to 24 degrees then mites pests attack high on chilly, cotton ,sesame, jowar and jute crops.

Tree 3

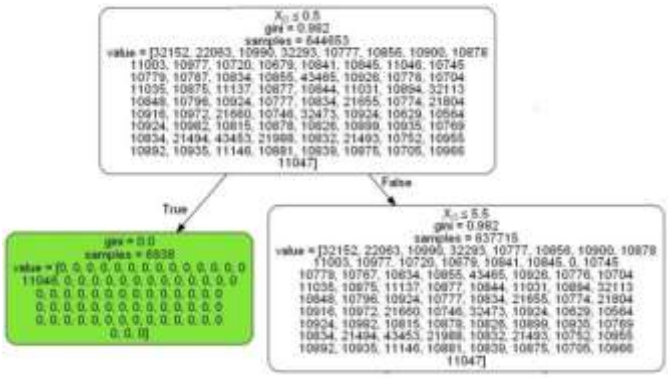

Fig 5. Tree based on Seasonal Occurrence

Rule 1: If it is dry season then maruc pod borer pest attacks on black gram crop.

Rule 2: If rule 1 is true then spodopteraexigua pest attacks on bengal gram crop.

Rule 3: If rule 2 is false and if season is spring, maruc pod borer pest attacks on bengal gram crop.

Rule 4: If it is heavy rainy season then maruc pod borer pest attacks on bengal gram crop.

Rule 5: If temperature is less than equal to 27 degrees then leaf folder pest attacks soybean and rice crops.

Rule 6: If temperature is less than equal to 30 degrees then bihar hairy caterpillar pest attacks sesame crop.

Rule 7: If temperature is less than equal to 21 degrees then marcua pod borer pest attacks red gram crop.

Rule 8: If temperature is less than equal to 20degrees then mite pest attack jute and jowar crops.

Rule 9: If temperature is less than equal to 24 degrees then ear head caterpillars pests attacks ragi crop.

Rule 10: If temperature is less than equal to 23 degrees then mites pest attack sesame, chilly and cotton crops.

Rule 11: If it is warm season then stem borer pest attacks rice, jowar and ragi crops.

Tree 4

May – June 2020
ISSN: 0193-4120 Page No. 5776-5788

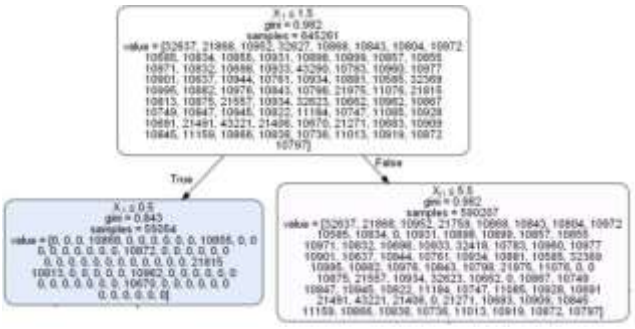

Fig 6. Tree based on Temperature

Rule 1: If temperature is less than equal to 21 degrees then maruc pod borer pest attacks high on black gram crop.

Rule 2: If rule 1 is true and if temperature is less than equal to 20 degrees then mite pests' attacks jowar and jute crops.

Rule 3: If rule 2 is false and if season is spring then ear head caterpillar pests attack high in ragi crop.

Rule 4: If it is dry season then stem borers pest will attack high on ragi crop.

Rule 5: If temperature is less than equal to 27 degrees then maruc pod borer pest attack high in black gram crop.

Rule 6: If temperature is less than equal to 23 degrees then leaf folder pest attack high in soybean and rice.

Rule 7: If temperature is less than equal to 36 degrees then ear head caterpillars pests attacks high in ragi crop.

Rule 8: If temperature is less than equal to 26 degrees then stem borer pest attacks rice.

Rule 9: If temperature is less than equal to 31 degrees then seedling blight pest attacks on jute crop.

Rule 10: If temperature is less than equal to 35 degrees then leaf hopper pest attacks on cotton crop.

Rule 11: If temperature is less than equal to 24 degrees then mites' pest attacks cotton, sesame and chilly crops.

Tree 5

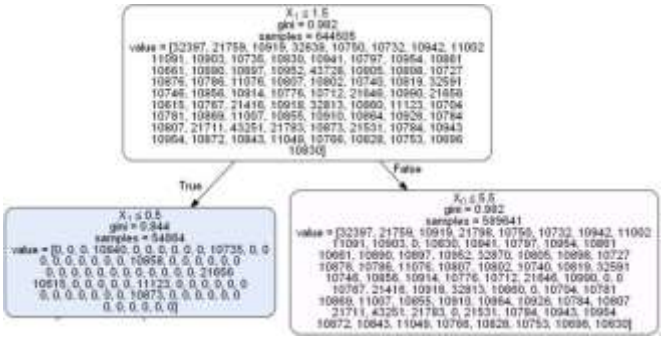

Fig 7. Tree based on Temperature

Rule 1: If temperature is less than equal to 21 degrees then maruc pod borer pest attack in black gram crop.

Rule 2: If rule 1 is true and if temperature is less than equal to 20 degrees then mite pest attack on jute and jowar.

Rule 3: If rule 1 is false and it is spring season then ear head caterpillars pest attack in ragi crop.

Rule 4: If temperature is less than 7 degrees then maruc pod borer pest attacks in black gram crop.

Rule 5: If temperature is less than 25 degrees then leaf folder pests attack soybean and rice crop.

Rule 6: If it is dry season then maruc pod borer pest attacks black gram crop.

Rule 7: If it is warm season then seedling blight pest attacks jute crop.

Rule 8: If temperature is less than 31 degrees then seedling blight pest attack jute crop.

Rule 9: If temperature is less than 35 degrees then marcua pod borer pest attacks red gram crop.

Rule 10: If temperature is less than 24 degrees then mites pest attack sesame, chilly and cotton crops.

May – June 2020
ISSN: 0193-4120 Page No. 5776-5788

Tree 6

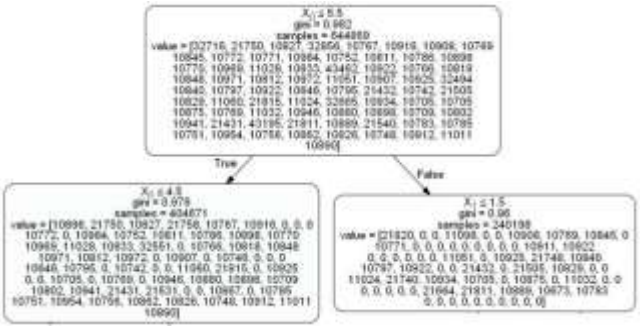

Fig 8. Tree based on Seasonal Occurrence

Rule 1: If it is a spring season maruc pod borer pest attacks black gram crop.

Rule 2: If rule 1 is true and if it is rainy season then maruc pod borer pest attacks high in black gram crop.

Rule 3: If rule 1 is false and if temperature is less than equal to 21 degrees then stem borer pest attacks rice, jowar and ragi crops.

Rule 4: If temperature is less than equal to 31 degrees then maruc pod borer pest attacks black gram crop.

Rule 5: If temperature is less than equal to 25 degrees then stem weevil pest attacks on jute crop.

Rule 6: If it is a dry season maruc pod borer pest attacks black gram crop.

Rule 7: If temperature is less than equal to 20 degrees then mite pest attacks jowar and jute crop.

Rule 8: If temperature is less than equal to 35 degrees then stem borer pest attacks jowar and rice crops.

Rule 9: If temperature is less than equal to 23 degrees then mites' pest attacks sesame, chilly and cotton crop.

Tree 7

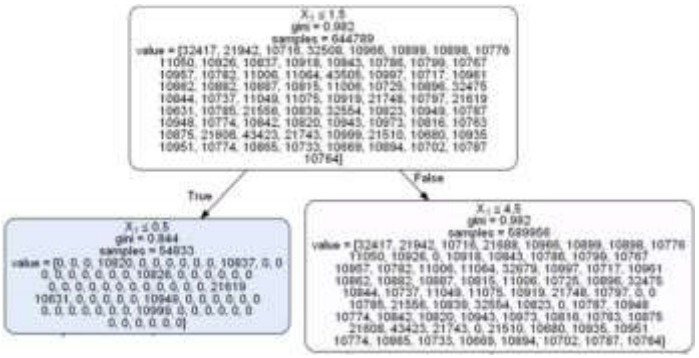

Fig 9. Tree based on Temperature

Rule 1: If temperature is less than equal to 21 degrees then maruc pod borer pest attacks black gram crop.

Rule 2: If rule 1 is true and if temperature is less than equal to 20 degrees then mite pest attacks jowar and jute crop.

Rule 3: If rule 2 is false and if temperature is less than equal to 24 degrees then ear head caterpillars' pest attacks in ragi crop.

Rule 4: If it is spring season then mites' pest attacks sesame, chilly and cotton crops.

Rule 5: If temperature is less than 23 degrees mites pest attacks sesame, chilly and cotton crop.

Rule 6: If it is spring season then aphids pest attacks black gram, cotton and chilly crops.

Rule 8: If temperature is less than 35 degrees then maruc pod borer pest attacks black gram crop.

Rule 9: If temperature is less than 29 degrees then ear head caterpillars pest attacks ragi crop.

Rule 10: If it is a spring season then marcua pod borer pest attacks red gram crop.

Tree 8

May – June 2020
ISSN: 0193-4120 Page No. 5776-5788

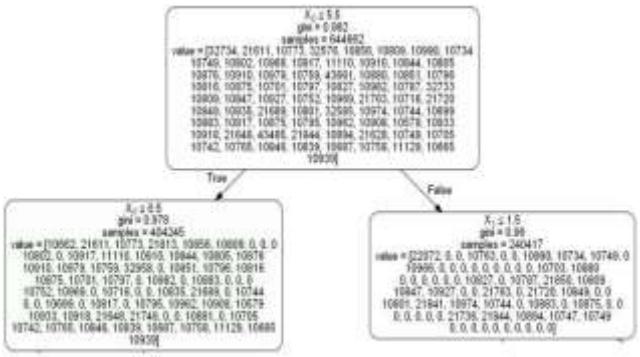

Fig 10. Tree based on Seasonal Occurrence

Rule 1: If it is spring season then maruc pod borer pest attacks on black gram crop.

Rule 2: If the rule 1 is true and if it dry season then maruc pod borer pest attacks on black gram crop.

Rule 3: If rule 1 is false and if temperature is 21degrees then stem borer pest attacks rice and jowar crops.

Rule 4: If it is rainy season then maruc pod borer pest attacks on black gram crop.

Rule 5: If temperature is less than equal to 30 degrees then maruc pod borer pest attacks on black gram crop

Rule 6: If temperature is less than equal to 23 degrees then stem weevil pest attacks on jute crop.

Rule 7: If temperature is less than equal to 20 degrees then mite pest attacks jowar and jute crop.

Rule 8: If it is a warm season then stem borer pest attacks rice and jowar crops.

Rule 9: If temperature is less than equal to 31 degrees then seedling blight pest attacks jute crop.

Rule 10: If temperature is less than equal to 35 degrees then stem borer pest attacks rice and jowar crops.

Rule 11: If temperature is less than equal to 24 degrees then mites' pest attacks cotton, sesame and chilly crops.

Tree 9

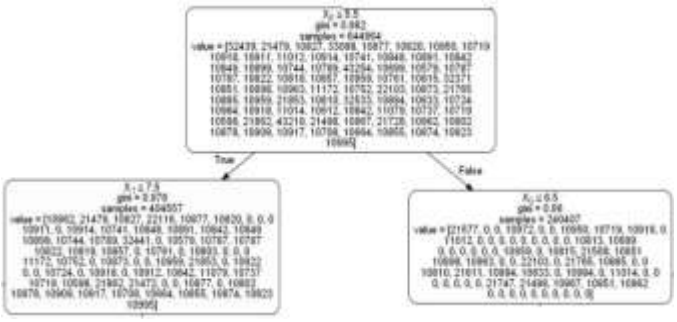

Fig 11. Tree based on Seasonal Occurrence

Rule 1: If it is spring season then maruc pod borer pest attacks in black gram crop.

Rule 2: If rule 1 is true and temperature is less than equal to 27 degrees then maruc pod borer pest attacks in black gram crop.

Rule 3: If rule 1 is false and if season is warm then ear head bug pest attacks jowar crop.

Rule 4: If temperature is less than equal to 36 degrees then stem borers pest attacks ragi crop.

Rule 5: If temperature is less than equal to 26 degrees then stem weevil pest attacks jute crop.

Rule 6: If temperature is less than equal to 24 degrees then panicle worm pest attacks jowar crop.

Rule 7: If it is dry season then stem borers pest attacks ragi crop.

Rule 8: If temperature is less than equal to 31 degrees then seedling blight pest attacks jute crop.

Rule 9: If temperature is less than equal to 29 degrees then army worms pest attacks on korra and ragi crops.

Rule 10: If temperature is less than equal to 35 degrees then leaf hopper pest attacks on cotton crop.

Rule 11: If temperature is less than equal to 21 degrees then mite pest attacks on jowar and jute crop.

May – June 2020
ISSN: 0193-4120 Page No. 5776-5788

Rule 12: If temperature is less than equal to 24 degrees then mites' pest attacks on chilly, cotton and sesame crops.

The trees generated give the classification of pests which occurred in unusual seasons and in different temperatures. Some pests like maruc pod borer, mites and stem borer effects the crops mostly in all seasons.

VI. CONCLUSION

Agriculture is a main occupation in Andhra Pradesh. Many crops are cultivated in this state, and pest attack is one of the biggest problems in agriculture. The problem is occurred due to climate, season and type of crop. Here for study, crop pest classification is done based on temperature and seasons. The data has twenty major crops with their pest name, temperature, and seasonal occurrence. The data set is in uncountable from so here MongoDB is used for data storage and retrieval using python programming language. Then machine learning is applied to dataset for prediction by using methodsDecision Tree (DT), Random Forest Classifier (RFC), K-Nearest Neighbour (K-NN), Gaussian Naïve Bays (GNB), Support Vector Machine (SVM) and Ada Boost Classifier (ABC) are applied. Among them Random forest classifier is proved to be best method for analysing the crop. By analysis RFC performs well for classification of various types of pest attack under different weather conditions. This classification assists the department of agriculture to take necessary decisions to control pests.

REFERENCES

[1] Anwesha Borthakur, PardeepSingh.(2012). Agricultural Research In India: An Exploratory Study. International Journal of Social Science & Interdisciplinary Research,Vol.1 Issue 9,ISSN 2277-3630.

[2] Agriculture Research Institute.(2009). Annual report of entomology work, Hyderabad, India.

[3] Ryan, H.,L.,P., Li-Minn Ang, Kah Phooi seng, J.C.Broster, Pratley J.E.(2018). Big data and machine Learning for crop protection, volume 151.

[4] Kotsiantis, S.,B (2007).Supervised Machine Learning: A Review of Classification Techniques, Informatica 31, 249-268.

[5] STATE-SPECIFIC INTERNVENTIONS FOR HIGHER AGRICULTURAL GROWTH https://icar.org.in/files/state-specific/chapter/25.htm

[6] HariGanesh S, Pritty Cindrella D., Joy Christy A..(2015).A Review On Classification Techniques Over Agricultural Data,International Journal Of Computer Science And Mobile Computing,Vol.4,Issue.5,Pg.49-45.

[7] Latha Umasankar, Velmurugan T.(2016). Analyzing Agricultural Text Data Using Classification Algorithms, The 3rd International Conference On Small & Medium Business.

[8] Hetal Patel P , Dharmendra Patel.(2016). A Comparative Study On Various Data Mining Algorithms With Special Reference To Crop Yield Classification. Indian Journal Of Science and Technology, Vol9(22), doi: 10.17485/Ijst/2016/V9i22/92713.

[9] Revathy, R., Lawrence, R.(2017).Classifying Crops Pest Data Using C4.5 Algorithm.IEEE International Conference On Intelligent Techniques In Control, Optimization And Signal Processing(Incos), doi:10.1109/Itcosp.2017.8303122.

[10] Umair Ayub.(2018). Predicting crop diseases using data mining approaches: Classification, IEEE Conference.1st international conference on Power, Energy and Smart Grid **, doi:** 10.1109/ICPESG.2018.8384523.

[11] Chandra Sekhar Ch, Uday Kumar J, Kishore Kumar B, Sekhar Ch.(2018). Effective use of Big Data Analytics in Crop planning to increase Agriculture Production in India. International Journal of Advanced Science and Technology, vol.113(2018).pp.31-40.

[12] Sagar B.M., Cauvery N.K. (2018). Agriculture Data Analytics In Crop Yield Estimation: A Critical Review, Indonesian Journal Of Electrical Engineering And Computer Science, Vol 12,Pp 1087-1093,Issn:2502-4752.

[13] Liakos K, Busato P, Moshou D, Pearson S, Bochtis D. (2018). Machine Learning In Agriculture : A Review, Us National Library Of Medicine National Institute Of Health, doi: 10.3390/S18082674.

May – June 2020
ISSN: 0193-4120 Page No. 5776-5788

[14] Ansif Arooj,Mohsin Riaz, Malik Naeem Akram.(2018). Evaluation Of Predictive Data Mining Algorithms In Soil Data Classification For Optimized Crop Recommendation, International Conference On Advancements In Computational Sciences.

[15] Avinash Kumar, Sobhangi Sarkar, Chittaranjan Pradhan.(2019). "Recommendation System For Crop Identification And Pest Control Technique In Agriculture", International Conference On Communication And Signal Processing (Iccsp), Doi:10.1109/Iccsp.2019.8698099.

[16] Mukesh Nagar, Mayura Nagar.(2017).Big data analytics in agriculture and distribution channel. IEEE Xplore, International Conference on Computing Methodologies and Communicaiton,doi:10.1109/ICCMC.2017.8282714.

[17] Lokesh Kumar, Shalini Rajawat, Krati Joshi.(2015).Comparative Analysis of NoSQL(MongoDB) with MySQL Database.International Journal of Modern Trends in Engineering and Research,ISSN:2349-9745.

[18] Amy Genender, J.D.,Feltheimer,M.A.(2018).Visualizing High Dimensional and Big Data, ELSEVIER, Procedia Computer Science, volume 140.

[19] KishanSingh Rajput, Bhavesh A.Oza.(2017).A Comparative Study of Classification Techniques in Data Mining",International Journal of Creative Research Thoughts,Volume:5,Issue:3,ISSN:2320-2882.

[20] Ayon Dey(2016).Machine Learning Algorithms:A Review.International Journal of Computer Science and Information Technologies,Vol.7(3),ISSN:0975-9646.

[21] Bhaskar Patel,N., Satish, G., Prajapati , Dr. Kamaljit I., Lakhtaria.(2012).Efficient Classification of Data Using Decision Tree, ISSN 2277 – 5048.

[22] Lei Shie, Qiguo Duan, Xinming Ma, Mei Weng(2012).The Research of Support Vector Machine in Agricultural Data Classification, IFIP International Federation for Information Processing.

[23] Gayathri,B.,M., Sumathi,C.,P.(2016).An Automated Technique using Gaussian Naïve Bayes Classifier to Classify Breast Cancer. International Journal of Computer Applications, Volume -148-No.6,2016.

[24] Sadegh Bafandeh Imandoust, Mohammad Bolandraftar.(2013).Application of K-Nearest Neighbour (KNN) Approach for Predicting Economic Events: Theoretical Background, International Journal for Engineering Research and Applications, Vol. 3, Issue 5.

[25] Jasmina,D., Novakovic, Alempije Veljovic, Sinisa, S., Ilic, MilosPapic.(2016).Experimental Study Of Using The K-Nearest Neighbour Classifier With Filter Methods.Research Gate.

[26] Arnu Pretorius, Surette Bierman, Sarel Steel, J.(2016). A Meta Analysis of Research in Random Forests for Classification. International Conference(PRASA-RobMech) Stellenbosch, South Africa.

[27] Abellán, J.(2013).Ensemble of decision tree based on imprecise probabilities an uncertainty measures", Information Fusion, 14,423-430.

[28] Xu, J., Li, H.(2007). AdaRank: A boosting algorithm for information retrieval, SIGIR.

[29] Freund, Y., Schapire, R.,E.,(1996).Experiments with a New Boosting Algorithm . MachineLearning,13th International, New Conference.

[30] Schapire, R.,T.(2003).The Boosting Approach to Machine Learning, An Overview. Nonlinear Estimation and Classification, Springer, New York.