INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

A STUDY AND ANALYSIS OF RICE AND WHEAT PRODUCTION IN INDIA BASED ON SEASON USING MACHINE LEARNING CLASSIFICATION ALGORITHMS

Ms. K. S. Leelavathi [1], Dr. M. Rajasenathipathi [2]

1. Research Scholar, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, Tamilnadu .
E-Mail: leelakarthi@gmail.com, Mobile: 9788341395.

2. Head of the Department, Department of Computer Technology, Nallamuthu Gounder Mahalingam College, Pollachi, Tamilnadu.

ABSTRACT:

Indian agriculture is based on season and rainfall. Most of the crops are cultivated in Kharif and Rabi season. Nowadays drastic changes in climatic conditions and rain fall leads to the biggest threat for the crop cultivation and gain more yield to fulfil the requirement of increasing population. Most of the peoples need rice and wheat grains as their daily meal. So better and advanced technology required to analyse and find the strategies for crop cultivation, protecting the crop from diseases and maximizing the crop production. Machine Learning techniques played a vital role in analysis and management of crop, soil, season, rainfall, crop diseases and production. This work attempts to find the efficient tree based ML classifier to classify the rice and wheat production data based on season. Random Forest, Random Tree, RepTree, NB Tree and J48 classifiers are used to classify the rice and wheat production data. The experimental results shows Random Forest classifier produces high accuracy value (98.3796%) in classification of crop production data based on season.

Keywords: Crop Production, Tree based classifiers, Machine Learning, Season

1. INTRODUCTION

In India, Rice and Wheat grains occupies major part in all kind of socio economic citizens food. Need of rice and wheat grains have been increasing day by day in large quantity (metric tones). Among all states of our country, only few states have major terrain, soil type, weather and climatic conditions to cultivate the paddy and wheat. West Bengal, Uttar Pradesh, Punjab, Andhra Pradesh,

Orissa, Telangana and Tamilnadu producing nearly 70% from the total production in every year. Uttar Pradesh, Madhya Pradesh, Punjab, Haryana and Rajasthan producing nearly 80 % of wheat from the total production in every year.

Majority of the crops cultivated in Kharif (South West (SW) monsoon) and Rabi (winter - end of SW monsoon) season in India. Kharif season (June to October) used to cultivate the monsoon based crops and Rabi season (October to March) used to cultivate the winter season based crops. In these two seasons, rice is mostly cultivated in kharif and wheat is mostly cultivated in rabi.

The various factors are considered to maximize the yield of crops such as soil type, fertility of the soil, season, climatic conditions, and irrigation type, disease affected in the crop and climatic conditions at the time of harvesting. Seasonal based cropping needs more attention and latest techniques to identify the suitable soil, irrigation and fertility of the soil to increase the yield and also protect the crops from the diseases. Nowadays Machine and Deep learning with IOT techniques played huge role in the agriculture.

Management of soil, water and crop, detection of crop diseases and yield prediction are required Machine Learning (ML) and deep learning techniques and strategies to improve the crop yield and maximize the profit in agriculture. From the various supervised and unsupervised ML models, Tree based supervised models effectively classify the table or spatial based data sets in the form of tree structure to determine the class variables based on features.

II LITERATURE SURVEY

Supervised and Unsupervised Machine Learning techniques and methods are applied to analyse, manage and predict the agriculture yield, soil fertility, season, crop diseases and production.

Ramesh et al, [2] attempts to find out the suitable model for yield prediction process. The author utilized k-Means clustering and Multiple linear regression method for the 44 years (196 to 2009) rain fall data of East Godavari district. K-means adopted to cluster the rain fall data into four clusters and Multiple linear regression used to find the association between the rainfall with area, production and year of rain fall.

L Brieman [3], analysing pros and cons of random forest algorithm. It generates more decision trees on dissimilar data samples and combines all decision trees for regression, classification and prediction systems. Inserting the deserved randomness helps to reduce the correlation and increase the strength.

Kumar et al [4] proposed Random forest based crop yield prediction model with high accuracy using temperature, soil fertility (ph value), rain fall, humidity and crop details based data set of past crop production details from various districts of India.

Suresh.G. et al, [5] utilized SVM to find the specific crop in crop and location data set with high accuracy rate. This system suggests required soil nutrients (NPK and Ph) values for crops like grams, rice, carrot and maize.

III. SYSTEM ARCHITECTURE

Architecture of tree based classification model from ML for rice and wheat production based on seasons is presented in fig.1. Random Forest, Random Tree, RepTree, NB Tree and J48 are used to classify the rice and wheat production data based on seasons.

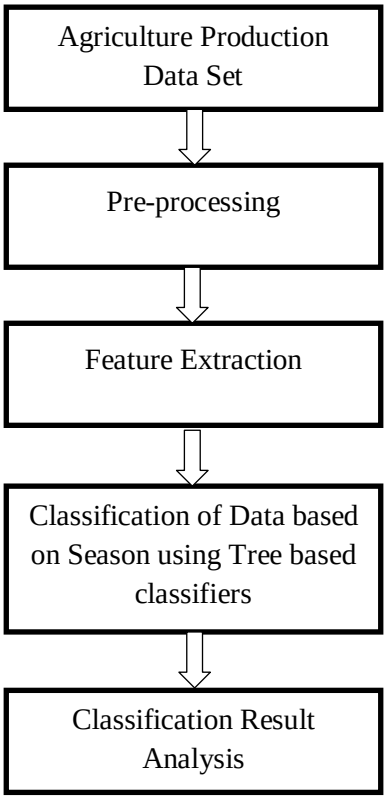


Fig.1. Architecture of Tree based classification model

Data set:

Agri production details data set is collected from the data.gov.in portal with state and district name, crop, year, season, area and production attributes. 18 years (1997 to 2014) crop production details are collected from the source [1].

Pre-Processing and Feature Extraction

The collected data set consists of 246091 instances with all vegetables, fruits, pulses, grains, etc crop production details. From the collected data set, rice and wheat production details are extracted in the pre-processing stage. Majorly rice and wheat production states data is separated and created a new data set for the classification process. The features state name, district name, season, crop, area and production are extracted with 8208 instances in the feature extraction phase. From the extracted features, season (nominal) is assigned as class attribute for classification.

Tree based classifiers from ML

- Random Forest: Ensemble and Bagging ML classifier model; It generates more decision trees on dissimilar data samples and combines all decision trees for regression, classification and prediction systems.
- Random Tree: Supervised ML classifier. Most frequently occurred tree from random forest and that particular tree applied to whole data set.
- RepTree : Reduced Error Pruning Tree generates continues (for regression) and discrete (for classification) outcome using generated decision trees.
- NB Tree: Combined version of Naïve Bayes classifier and Decision Tree to provide the probabilities of outcome.
- J48: Statistical classifier builds the decision tree by using SubTree raising and reduced error pruning methods.

Performance Analysis

The measures are taken to analyse various tree based ML classifiers on rice and wheat production data classification based on season are classification accuracy, time consumed to build the model, true positive, false positive, precision, recall and f-measure.

From the result of the classification model, correctly identified the positive class is called True Positive (TP), correctly identified negative class is called True Negative (TN). Incorrectly identified positive class is called False Positive (FP) and incorrectly identified negative class is called False Negative (FN). Sum of correctly classified instances (TP and TN) is divided by total instances returns the percentage of correctly classified instances.

Precision = TP / sum of TP and FP

Recall = TP/ sum of TP and FN

F-Measure = 2 *((TP * FP) / (TP+FP))

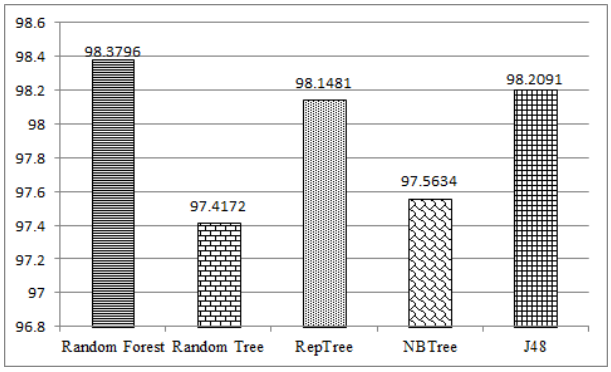
IV. RESULTS AND DISCUSSION

Classification accuracy of rice and wheat production data set using five tree based ML classifiers are showed in the table 1. All five tree based ML classifiers performs well in the season based classification process.

Table 1. Classification Accuracy

Classifier	Correctly classified Instances	Time taken to build (Seconds)
Random Forest	98.3796	1.48
Random Tree	97.4172	0.03
RepTree	98.1481	0.11
NBTree	97.5634	2.73
J48	98.2091	0.11

Fig.1. Classification Accuracy



Random forest produced high accuracy value than other classifiers such as J48, RepTree, NB Tree and Random Tree. From the 8208 instances, 8075 instances (98.3796 %) are correctly classified and 133 instances (1.6204 %) are incorrectly classified by the Random forest classifier with consumption of 1.48 seconds to build the model. Random Tree classifier takes minimum time (0.003 seconds) to build the model but produced low accuracy (97.4172%) than other classifiers. Naïve Bayes based tree takes more time (2.73seconds) to build the model.

Table 2. Performance Measures

Classifier	TP	FP	Precision	Recall	F-Measure
Random Forest	0.984	0.016	0.984	0.984	0.984
Random Tree	0.974	0.026	0.974	0.974	0.974
RepTree	0.981	0.018	0.982	0.981	0.981
NBTree	0.976	0.024	0.976	0.976	0.976
J48	0.982	0.018	0.982	0.982	0.982

The above table 2, shown that the TP, Precision and Recall values of Random Forest are 0.984 and FP value is 0.016. It reveals that the Random forest shown good accuracy than J48, RepTree, NBTree and Random Tree.

Fig.2. TP and FP Values of Tree based Classifiers

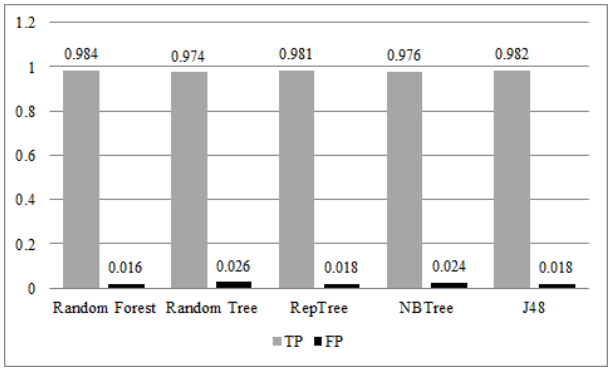
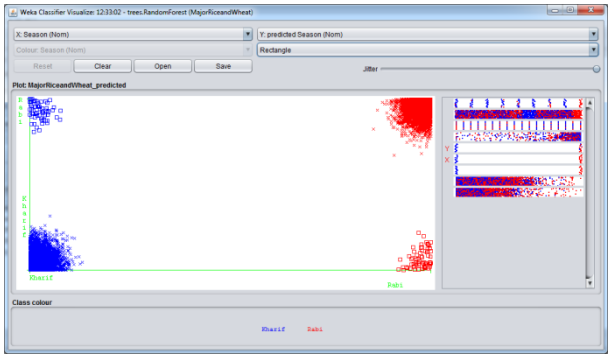


Fig.3. Actual and Predicted seasons



The actual and predicted seasons of rice and wheat production details by the Random forest classifier is shown in the fig.3.

V. CONCLUSION AND FUTURE SCOPE

This study shows that Random Forest classifier from ML achieves good results with high accuracy than other tree based ML classifiers in the rice and wheat production details data set based on the season. Random forest is suitable classifier to classify all variety of crop production details based on the season. In future, more parameters to be included like soil type and fertility, irrigation and water test results, rain fall details for all crops to classify and also predict the suitable crop for the particular soil in the region with good accuracy rate. To enhance the accuracy of Random Forest classifier from 98.3796 for additional parameters to be included with this data set, improved pre-processing and feature selection techniques will be added in future.

REFERENCES

1. https://www.data.gov.in
2. Ramesh D and Vishnu V B., Data mining techniques and applications to agricultural yield data. In: Intl. Jounrl of Adv. Research in Comp. and Comm. Engineering.2013; 2(9).
3. L Brieman , Random Forest, Machine Learning, Kluwer Academic Publ. 2001, 45, 5-32
4. Kumar, Y J Nagendra, Spandana, Vaishnavi, Neha Devi., "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector." In 2020 5th Intl.Conf. on Comm. and Electronics Sys. (ICCES), IEEE, 2020.
5. Suresh.G, A.S.Kumar and S.Lekashri., "Efficient Crop Yield Recommendation Sys. Using ML For Digital Farming." Intl. Jrnl. of Modern Agri. Vol.,10, No., 2021 pp.906-914.
6. A Suresh, P G Kumar and M R Latha. "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN." In 2018 3rd Intl. Conf, on Comm. and Electronics Systems (ICCES), IEEE, 2018. pp. 88-93