

# Crop Production Prediction System

Shambhavi Dighe<sup>1</sup>, Anuja Jagdale<sup>2</sup>, Amar Chadchankar<sup>3</sup>

<sup>1,2</sup>Student, Department of Information Technology, ZCOER, Pune, India

<sup>3</sup>Professor, Department of Information Technology, ZCOER, Pune, India

**Abstract:** In agriculture field where farmers and agriculture businesses have to make countless decisions every day and difficult complexities involves the various factors influencing them. An essential concern for agricultural planning intention is the accurate yield estimation for the many crops involved in the planning. DM techniques are approach for accomplishing practical and useful solutions for this problem. Agriculture has been an obvious aim for big data. Environmental situation, variability in soil, input levels, combinations and goods prices have made it all the more related for farmers to use information and get help to make serious farming decisions. The difficulty of yield prediction is a major problem that remains to be solved based on available data. DM techniques are the better choices for this purpose. Different Data Mining techniques are used and evaluated in agriculture for estimating the future year's crop production. This paper presents a brief analysis of crop yield prediction using classification algorithm NB, SVM and KNN. These algorithms are used to classify the crop production, whether it will be low, average, high or very high.

**Keywords:** Agriculture; data mining; classification; grass grub; Naïve Bayes; Support Vector Machine, kth Nearest Neighbor.

## 1. Introduction

Agriculture is the spine of Indian Economy. In India, many of the farmers are not getting the expected crop yield due to several reasons. The agricultural yield is primarily depends on weather conditions. We are considering the grass grub dataset to predict crop production. In this context, the farmers necessarily requires a timely counsel to predict the future crop productivity and an analysis is to be made in order to help the farmers to maximize the crop production in their crops. Yield prediction is an important agricultural problem. Every farmer is involved in knowing, how much yield he/she is about expect. In the past, crop prediction was performed by considering farmer's previous history on a particular crop. The amount of data is large in Indian agriculture. The data when become information is highly useful for many purposes. Data Mining is widely applied to agricultural problems. Data Mining is used to analyze large data sets and establish useful classifications and patterns in the data sets. The overall goal of the Data Mining process is to extract the information from a data set and transform it into understandable structure for further use. In this paper the main aim is to create a user friendly interface for farmers, which gives the analysis of rice production based on available data. Different Data mining techniques were used to predict the crop yield for maximizing the crop productivity

## 2. Literature survey

In [8] author articulated that different classifiers can be used on different features in ensemble model for predicting true class. Each model for ensemble classifier trained on different set of features to achieve the best results. And high vote getting class will be allocated to test sample. In [9] authors provide an extensive survey on classifiers and explain the ways for better prediction. Authors explained that we cannot solve all the problems with same classifier. Different classifiers work in different way so different classifiers can be used in different scenarios. In following papers [2-3] NN is used as classifier for predicting the diseases in fruits. From the research article [3], the researcher expresses that large amount of data which is collected and stored for analysis. Making appropriate use of these data often leads to considerable gains in efficiency and therefore economic advantages. There are several applications of Data Mining techniques in the field of agriculture. The researchers implemented [4] KMeans algorithm to forecast the pollution in the atmosphere, the K Nearest Neighbour is applied [10] for simulating daily precipitations and other weather variables and different possible changes of the weather scenarios are analyzed [7] using Support Vector Machines. Soil profile descriptions were proposed [6] by the researcher for classifying soils in combination with GPS based technologies. They were applied K-Means approach for the soil classification. In a similar approach, crop classifications using hyper spectral data was carried out [1] by adopting one of the data mining approach i.e. Support Vector Machines. One of the researcher used [9] an intensified fuzzy cluster analysis for classifying plants, soil and residue regions of interest from GPS based colour images. In the agricultural science, clustering techniques are found in grading [5] apples before marketing. Weeds were detected on precision agriculture [13]. The researchers worked [8] on rainfall variability analysis and its impact on crop productivity. The effect of observed seasonal climatic conditions such as rainfall and temperature variability on crop yield prediction was considered [7] through an empirical crop model. Furthermore, there are two approaches to investigate the impact of climate change on crop production which include the crop suitability approach and the production function approach [6]. Researchers were found that the yields of winter wheat are reduced when temperatures rise, due to the consequent reduction of the growth phases of the plant [2] and also concluded that the complexity of a model was based on the level

of detailed analysis [11] or it was less detailed with only estimations of moisture content [2].

#### A. Problem Statement

- Agriculture field is one of the most important parts in the economical growth of the country. Very extensive work is going on from last decade to improve the productivity by using data mining & evolutionary techniques.
- To produce enough food this field should be free from the disease on the crops. So this is the one of the major issue which make farmers to produce enough food.
- There is a limited analysis performed for loss of crops, due to disease or growth of insects. The correct prediction of loss in crops helps farmer to apply some suitable action to overcome the issues and maintain its expected production.

#### B. Design goals

The objective of this project is to predict the grass grub damages and to identify the best classification technique to predict the grass grub damages.

#### C. Proposed work

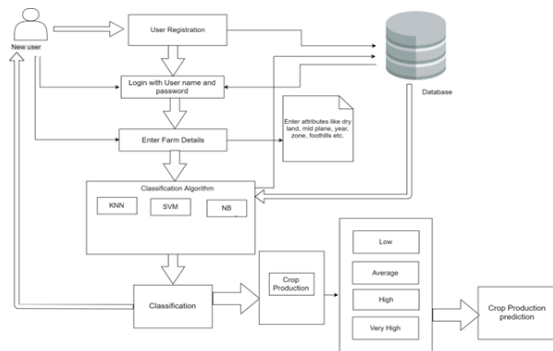


Fig. 1. The Proposed System

In the proposed system, crop production prediction is performed. Grass grub is an insect that cause diseases in plants. This insect can destroy the crops and due to this the enough food can't be generated. To predict the loss from population of grass grub proposed system used different data mining techniques. The dataset used in this system is the grass grub dataset which is having 8 features (attributes). There are four classes named as: low, average, high and very-high in the dataset. In preprocessing, generate the binary version of dataset

which has only two classes named as: low (combination of low class and average class) and high (combination of high class and very-high class). In rest of the paper, two class dataset is referred to as binary data and four class data is referred to as original data. 70% of the data has been used for model training; rest of the data has been used for testing the model performance on different criteria.

### 3. Conclusion

The different data mining techniques for classification are used SVM, KNN and NB. And the accuracies of each data mining techniques are found. According to this we can say that prediction is performed to reduce the loss in production.

### References

- [1] A. Petrakova, M. Affenzeller and G. Merkurjeva, "Heterogeneous versus homogeneous machine learning ensembles," *Information Technology and Management Science*, vol. 21, pp. 135-140, 2015.
- [2] Y. H. Kim et al., "Crop pests prediction method using regression and machine learning technology: Survey," in *International Conference on Agriculture and Biosystem Engineering*, 2008. Berlin: IERI Procedia, 2009, pp. 901-908.
- [3] M. Watts and S. Worner, "Predicting the distribution of fungal crop diseases from abiotic and biotic factors using multi-layer perceptrons," in *International Conference on Neural Information Processing*, 2008. Berlin: Springer, 2009, pp. 901-908.
- [4] Y. Chtioui, S. Panigrahi and L. Francl, "A generalized regression neural network and its application for leaf wetness prediction to forecast plant disease," in *Chemometrics and Intelligent Laboratory Systems*, 1999. USA: Elsevier, 1999, vol. 48, pp. 47-58.
- [5] D. Moshou et al., "Crop health condition monitoring based on the identification of biotic and abiotic stresses by using hierarchical selforganizing classifiers," in *Precision agriculture'15*, 2015. Wageningen Academic Publishers, 2015, vol. 8, pp. 619-626.
- [6] R. Pahlavan, M. Omid and A. Akram, "Energy input output analysis and application of artificial neural networks for predicting greenhouse basil production," *Energy*, vol. 37, pp. 171-176, 2011.
- [7] R. Pascanu, C. Gulcehre, K. Cho and Y. Bengio, "How to construct deep recurrent neural networks," in *ICLR*, 2014.
- [8] G. P. Zhang, "Neural Networks for Classification: A Survey," in *IEEE Transactions on Systems, Man and Cybernetics*, 2000. Atlanta: IEEE, vol. 4, pp. 451-462.
- [9] A. Mucherino, J. Papajorgji, and P. M. Pardalos, "K-nearest neighbor classification," in *Data Mining in Agriculture*, 2009. New York: Springer, 2009, vol. 4, pp.83-106.
- [10] P. A. Paul and G. P. Munkvold, "A model-based approach to preplanting risk assessment for gray leaf spot of maize," *Phytopathology journal*, vol. 94, pp. 1350-1357, 2004.
- [11] Patel, Hetal and D. Patel, "A brief survey of data mining techniques applied to agricultural data," *International Journal of Computer Applications*, vol. 9, pp. 6-8, 2014.
- [12] M. Stoces, J. Vanek, J. Masner and J. Pavlek, "Internet of things (iot) in agriculture-selected aspects," *Agris on-line Papers in Economics and Informatics*, vol. 8, pp. 83-88, 2016.
- [13] B. Milovic and V. Radojevic, "Application of data mining in agriculture," *Bulgarian Journal of Agricultural Science*, vol. 5, pp. 18- 21, 2017.