

# Crop yield forecasting and Fertilizer recommendation using Voting classifier

<sup>1</sup>B. Chakradhar, <sup>2</sup>D. Sai Veena, <sup>3</sup>Ch. Devasai Pranatheswar, <sup>4</sup>A. Vivek

<sup>1</sup>Student, <sup>2</sup>Student, <sup>3</sup>Student, <sup>4</sup>Student

<sup>1</sup>Computer Science and Engineering

<sup>1</sup>Lendi Institute of Engineering and Technology, Vizianagaram, India

**Abstract** - India being an agriculture country, its economy predominantly depends on agriculture yield growth and agroindustry products. Yield forecasting is a very important issue in agriculture. In order to predict the crop yield, all the attributes of data will be analyzed like temperature, type of soil, Nutrient value of the soil in that region, amount of rainfall in that region, soil composition can be determined. We train the data with various suitable machine learning algorithms like random forest and voting classifier for creating a model. The system comes with a model to be precise and accurate in forecasting crop yield and deliver proper recommendations about required fertilizer ratio based on atmospheric and soil parameters of the land which enhance to increase the crop yield and increase farmer revenue. Weather related information like rainfall, temperature and humidity used to forecast the better crop.

**Index Terms**-Yield Forecasting, Random Forest, Voting Classifier, Fertilizer Recommendation \_

## I. INTRODUCTION

Agriculture plays an essential part in an economy's life. They are the backbone of our country's economic system. One of the key problems confronting farmers is selecting the right crop for cultivation. Selection of crops is determined by several factors such as temperature, soil composition, market prices etc. Yield forecasting is a very important issue in agriculture. Any farmer is interested in knowing how much yield he is about to expect and what is the crop that is suitable for the land. Analyze the various related attributes like location, pH value from which alkalinity of the soil is determined. Along with it, the percentage of nutrients like Nitrogen (N), Phosphorous (P), and Potassium (K).

The voting classifier is an ensemble learning method that combines several base models to produce the final optimum solution. The base model can independently use different algorithms such as KNN, Random forests, Regression, etc., to predict individual outputs. This brings diversity in the output, thus called Heterogeneous ensembling. In contrast, if base models use the same algorithm to predict separate outcomes, this is called Homogeneous ensembling. The voting classifier is divided into hard voting and Soft voting.

Hard voting is also known as majority voting. The base model's classifiers are fed with the training data individually. The models predict the output class independent of each other. The output class is a class expected by the majority of the models.

In Soft voting, Classifiers or base models are fed with training data to predict the classes out of m possible courses. Each base model classifier independently assigns the probability of occurrence of each type. In the end, the average of the possibilities of each class is calculated, and the final output is the class having the highest probability.

The voting classifier aggregates the predicted class or predicted probability on basis of hard voting or soft voting. So, if we feed a variety of base models to the voting classifier it makes sure to resolve the error by any model.

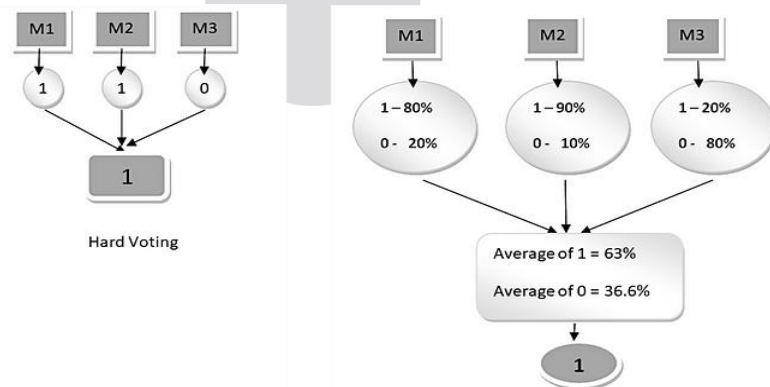


Fig.1. Left: Hard Voting, Right: Soft Voting

Scikit-learn does not provide implementation to compute the top-performing features for the voting classifier unlike other models, but I have come with a hack to compute the same. You can compute the feature importance by combining the importance score of each of the estimators based on the weights.

## II. LITERATURE SURVEY

**2.1 VIRENDRA PANPATIL ET :** It had accomplished gigantic work for Indian ranchers by making productive yield proposal framework. They created framework utilizing classifier models, for example, Decision Tree Classifier, KNN, and Naive Bayes Classifier. The proposed framework can be utilized to figure out best season of planting, development of plant and Plant reaping. They utilized distinctive classifier for accomplishing better exactness for instance: Decision tree shows less precision when dataset is having more varieties yet Naive Bayes gives preferable exactness over choice tree for such datasets. The best favorable position of framework that it can without much of a stretch versatile all things considered/be utilized to test on various yields.

**2.2 MAYANK ET:** It have presumed that this paper fabricate extemporized framework for crop yield utilizing administered AI calculations and with objective to give simple to utilize User Interface, increment the precision of crop yield forecast, investigate distinctive climatic boundaries, for example, overcast cover, precipitation, temperature, and so on In the proposed framework they zeroed in on MAHARASHTRA State for implantation and for information gathering they utilized govt. site, for example, [www.data.gov.in](http://www.data.gov.in). For crop yield forecast they utilized calculations, for example, Random Forest Algorithm and for convenience they created website page so it will be not difficult to use for all. The primary favorable position of proposed framework is precision rate is more than 75 percent on the whole the yields and areas chose in the examination.

**2.3 SHWETA ET:** It has inferred that this paper will survey those different utilizations of AI in the cultivating areas. And furthermore, helps in can be select appropriate crop select land and select season settled utilizing these procedures. The calculations use are Naïve Bayes and K-Nearest Neighbor. The calculations are utilizing precision of execution.

**2.4 AMIT KUMAR ET:** It has presumed that this paper helps in foreseeing crop arrangements and augmenting yield rates and making advantages to the ranchers. Additionally, Using Machine learning applications with farming in foreseeing crop sicknesses, examining crop copies, diverse water Page | 6 system designs. The calculations utilized are fake neural organizations. The serious issue with neural organization is that the proper organization which suits best for the arrangement is difficult to accomplish and it incorporates experimentation. The second issue with neural organization is the equipment reliance as the calculation incorporates more calculations in reverse and forward the preparing needs more. Assurance of appropriate organization structure requires insight and time. The proposed framework likewise centers around crop determination utilizing natural just as financial variables. The framework likewise utilizes the monetary factor that is the cost of the crop which assumes a significant part on the off chance that if the yields with same yield yet unique yield cost. The framework additionally utilizes other strategy which is crop sequencing which gives a full arrangement of yield which can be developed all through the season. The proposed framework likewise centers around crop choice utilizing ecological just as financial variables. The framework likewise utilizes the monetary factor that is the cost of the crop which assumes a significant part on the off chance that if the crops with same yield yet unique yield cost. The framework additionally utilizes other technique which is crop sequencing which gives a full arrangement of yield which can be developed all through the season.

## III. CONCLUSIONS

A farmer was curious about how much output he should expect during the rising season. Previously, this yield estimate was based on the farmer's long-term expertise with specific yields, crops, and meteorological circumstances. Instead of worrying about crop forecasts, farmers go straight for yield forecasting with the current technique. A forecasting model which may be used for future forecasting of crop yield.

## References

- [1]Jin Chaun, Qin Qiming, Zhu Lin, Nan Peng, Abduwasit Ghulam, "TVDI based Crop Yield forecasting model for Stressed Surface", international conference on Autonomous Region, 2007.
- [2]Joon-Goo Lee, Haedong Lee, Aekyung moon, "Segmentation method of COI. for Monitoring and forecasting of the crop growth",international conference on embedded system research section , 2014.
- [3]S. Das. Filters, wrappers and a boosting-based hybrid for feature selection. In International conference on machine learning 1 (2001), pp. 74–81.
- [4]H. Liu and Y. Lei, Toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng. 17 (2005), pp. 491–502.
- [5]Caruana, R., & Freitag, D. (1994). Greedy attribute selection. Proceedings of ICML-94.Drucker, H. (1997).
- [6]Fast committee machines for regression and classification. Proceedings of KDD-97.
- [7]Freund, Y., & Schapire, R. (1999). A short introduction to boosting. Journal of the Jap.
- [8]Soc. for AI. Hall, M. A. (1999). Correlation based feature selection for machine learning. Doctoral dissertation, The University of Waikato, Dept of Comp. Sci.
- [9]John, G. H., Kohavi, R., & Pfleger, K. (1994). Irrelevant features and the subset selection problem. Proceedings of ICML-94.

- [10]Kohavi, R. (1995). Wrappers for performance enhancement and oblivious decision graphs. Doctoral dissertation, Stanford University, Comp. Sci. Dept.
- [11]Koller, D., & Sahami, M. (1996). Toward optimal feature selection. Proceedings of ICML-96.

