International Journal of Current Trends in Engineering & Technology
www.ijctet.org, ISSN: 2395-3152
Volume: 07, Issue: 06 (November -December, 2021)

Crop Management and Forecasting Using Machine Learning

Ankur Ojha
Department of Mechanical
Engineering Galgotias College of Engineering and Technology, (AKTU),
Knowledge Park 2, Greater Noida
ankur15ojha@gmail.com

Ankit Ojha, Ajay Kumar, Anmol Dwivedi
Department of Information Technology
Galgotias College of Engineering and Technology (AKTU), Knowledge Park 2, Greater Noida
ankit2705ojha@gmail.com, ajay.2606@yahoo.com, anmoldwivedi@outlook.com

Abstract: - Agribusiness assumes a predominant part in developing the nation's economy. Climate and other ecological changes have become a significant danger in horticulture. AI (ML) is a fundamental methodology for accomplishing pragmatic and powerful answers for this issue. Crop Yield Prediction includes forecasting the crop yield from accessible, chronicled accessible information like climate parameters, soil boundary and noteworthy crop yield. This paper focused on anticipating the harvest yield dependent on the current information by utilizing K-means clustering and Random Forest calculation. Genuine information of data of India was utilized for building the models, and the models were tried with samples. The forecast will serve the farmer to foresee the harvest yield before developing onto the farming field. To foresee the harvest yield in future, precisely the k-means clustering and Random Forest are the generally excellent and well-known algorithms and directed AI calculation is utilized.

Keywords: Yield Analysis; Product Yield, Machine learning; Result Prediction; Random Forest Algorithm, decision tree Algorithm, k-means clustering, ensemble approach.

1. Introduction

Farming and agriculture are the firmness of every nation. In the farms like India, which produces an unexpected rising demand for agriculture products and diet due to a high number of people, approaches in the farming area are expected to fit the necessities. From an early age, farming is deemed the foremost and the leading lifestyle followed in nations likewise India. Aged personalities grow the herbs on their farm; hence people have been served requirements of self. Accordingly, the actual yields were grown and used by many living- things like people, creatures, fowl, cows, and seabirds.

The healthy products grew on the farm which the human's heads have taken to a good and healthy lifestyle. After discovering the latest creative tools and procedures, the farming area gradually diminishes. Thus specific, with immense creativity, humans continue to focus on cultivating synthetic food items that are composite crops, and they head to an injurious lifestyle, and it's good for human beings. In today's era, ordinary human beings do not possess knowledge regarding improving the products in the best interest and position. That's why specific cropping methods and procedures and periodical weather circumstances are also moving against the underlying treasures like clay, water, and atmosphere, which will head to the vulnerability of feeding food. By examining the culmination and problems like climate, warmth, and many other circumstances, we have understood. In India, we are in this situation, and we have no decent process and tools & techniques to succeed the circumstances struggled. We researched various ways to enhance the financial increase in farming and horticulture. You can observe various techniques to enhance and advance the farm crop and the condition of the products. Machine learning is beneficial for forecasting farm yield generation. Usually, Machine learning is the subject of programming algorithms that try to adjust to circumstances automatically through experience and according to data. Machine learning tools and programs recognize users to examine data from many different perspectives or viewpoints, classify, and shorten the associations distinguished. If we talk more about machine learning, it is the process of finding relationships and behaviour amongst hundreds of levels in prominent structured and unstructured datasets available in plenty of amounts. The models, organizations, or associations in all this information can give the report. Raw data can be converted into structured data from the data set using clustering algorithms like k-means, and we can apply different classification algorithms to classify the result. We can see traditional behaviour and anticipated biases.

Like, review data about product creation will assist the producers in recognizing the disasters of crops and blocking them in the eternity. Crop forecasting is a significant farming issue faced by each farmer. Indian farmers always think of knowing and gathering yield information to prepare themselves before growing the particular crop and profit from it. In the early days, crop forecasting was estimated by comparing farmers' prior knowledge of a singular product and its production. The Agricultural yield fundamentally turns on climate situations, bugs, soil

condition, moisture, a distance of market from the field, and preparation of crop cultivation. Specific knowledge around farm production is the foremost essential information for forming decisiveness compared to farm uncertainty about different parameters and management. Thus, this paper suggests an excellent concept to forecast crop cultivation in farmland. The Producer will monitor the yield of the product as per their requirements like water needed, fertilizers, temperatures, soil condition, and the size of the land according to the crop.

II. Literature Review

The purpose of this literature review is to study 10 years of research papers on crop yield prediction. Here we have investigated soft computing techniques and algorithms that are currently more efficient to solve our objective. [R1] SVM technique helps estimate soil and moisture, use remote sensing data. Each technique has its own merits and demerits, so we can get our desired results by combining more than one technique. K-means Clustering for grouping data, Random Forest regressors and decision tree regressors are the most efficient technique to obtain more accurate results from the data set. Getting climate, economic condition, and pollution data will predict a more satisfactory result. Machine Learning (ML) trades with obstacles wherever the connection in data and product variables is not identified or arduous to get. This "learning" word indicates the involuntary possession of structural information from samples of whatever is occurring in detail. Unlike conventional analytical techniques, ML seems not to create premises regarding the exact formation of the information model, which explains the information. This feature is beneficial for designing complicated non-linear operations, such as a crop yield prediction and management function. ML routines are every day vigorously applied to Crop Yield Prediction. An unsupervised machine algorithm such as the K-means algorithm is used to group the data based on similar behaviour in the crop prediction.

The supervised machine learning algorithm has separated input and output behaviour from each other, which must be foretold of a provided collection of predictors. Practising this set, we create a method to map or indicate desired output with the input parameters. The training method proceeds until the model obtains the aspired level of precision on the training dataset provided. Examples of Supervised Learning are the Linear regression Algorithm, random forest algorithm, Support vector machine, decision tree. Examples of Unsupervised Machine learning Algorithms are K-means for clustering algorithms in the Apriori algorithm.

2. Methodology

In the machine learning and analytic data field, the ensemble approach uses multiple learning algorithms to achieve much auspicious performance than could be achieved from any of the constituent learning algorithms solely. If elaborated more, we can say that the ensemble approach is a machine learning methodology that consolidates specific base models to produce one optimal and pre accurate predictive model with much efficiency. We will be using a two-level custom ensemble approach to solve this particular issue. The first level approach is an unsupervised machine learning algorithm to cluster data in different groups, and on the second level design, the approach is supervised machine learning algorithms.

Types of Learning

1. Unsupervised Machine Learning

Unsupervised machine learning algorithms understand patterns of a dataset externally reference to public or labelled results. Not like supervised machine learning, unsupervised machine learning techniques can't be immediately used to regression or a classification problem because the user have no notion what can be the outcome from the input and what output data might be, obtaining it impracticable for the user to train the algorithm using the idea you usually would try to do. Unsupervised learning can alternatively be utilized to determine the underlying and unobserved structure of the dataset. In other words, you can define unsupervised machine learning as the technique to labialize the unlabeled data. Ex: K-means Clustering, Apriori Algorithm, Principal Component Analysis, Singular Value Decomposition.

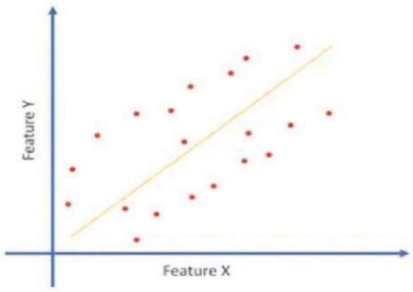


Fig.1. Unsupervised Machine Learning

1. Supervised Machine Learning Algorithm

Supervised learning is the numerous traditional sub-branches of machine learning in today's era. Usually, People who start with machine learning will commence their course with supervised learning algorithms of machine learning. Supervised machine learning algorithms are intended to learn case by case. The term "supervised" learning arises from the belief. You have somebody guru or teacher who will guide you throughout the process. When you train a supervised machine learning algorithm, the training data from the dataset must be comprised of levelled data of inputs paired with the accurate outputs set. The algorithm explores common patterns in the available data that correlate with the aspired output data throughout the training process. Later training, a supervised learning algorithm will be used in

different hidden data and will start determining which label the further input data will be categorized based on previously available training data. The purpose of a supervised learning model and algorithm is to forecast the fitting label for recent manifested input data from the user. In mathematical terms, a supervised learning algorithm shows as:

$$y = f(x)$$

Where Y is the forecasted output data found by a mapping function that selects a class to a particular input data x, the function has been used to correlate input features to a forecasted output. The machine learning model generates the result during the training of the dataset.

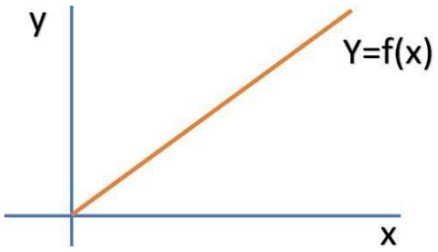


Fig.2.Result of the defined function

4. Outcomes

Tools Used

We have used python as the core of this project and for research work. Python libraries such as sci-kit-learn pandas and NumPy for prediction and calculation purposes. We have used a flask web framework, and for the front-end, we have used HTML, CSS and JavaScript to show the web page's result.

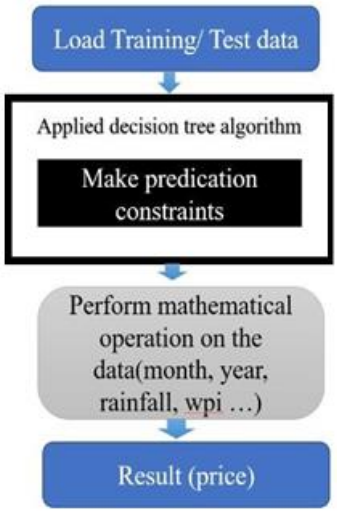


Fig.3. Decision Tree

Dataset Used

All datasets used in this research work are sourced from government site data.gov and open sources like Kaggle. Data was collected between 2012 and 2018 for 23 crops primarily cultivated in India. The dataset contains features like rainfall and WPI along their production month. The parameters that we have decided on for this study are listed below: the production's month, year of the production, rainfall during that particular period, and WPI (wholesale price index) for that particular period.

Decision Tree: **-** A decision tree is a graphical illustration of specific decision conditions applied when complicated branching befalls in a structured decision manner. A decision tree is an imminent model based on a branching range of Boolean tests that use detailed data to make more generalized outcomes. The main ingredients of a decision tree include decision points described by nodes, actions and distinct preferences from a decision point. Each precept within a decision tree is described by pursuing a series of paths from the root to the node to the next node until action is attained. First, we store all the CSV data files into a crop dictionary with the key name of the crop. We have stored the base price of all crops in the base dictionary and rainfall in the rainfall list for the current year.

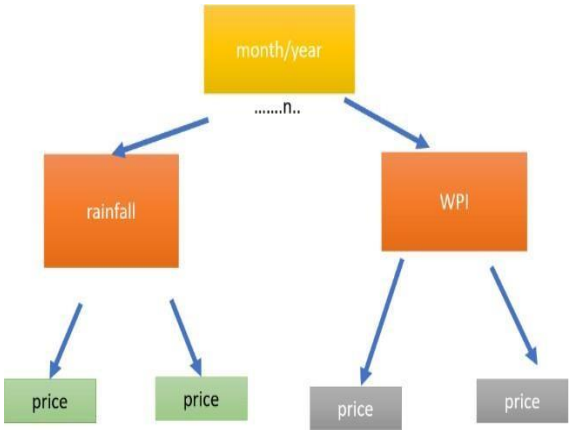
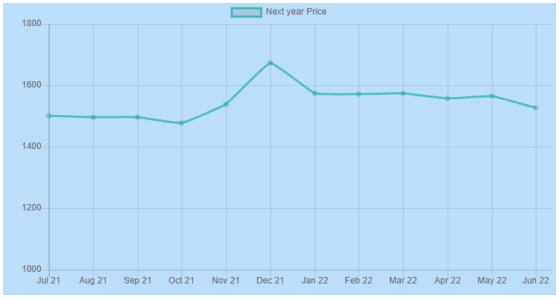


Fig.4. Example of a decision tree implementation



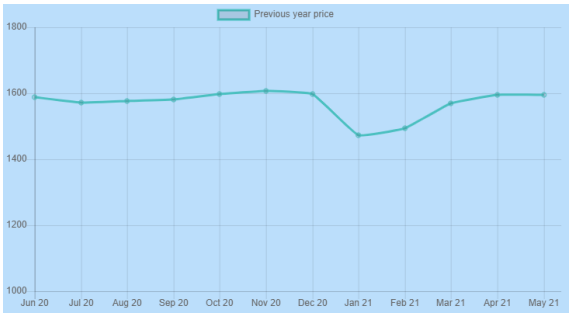
Next 12-month price prediction of wheat

Procedure: **-** We loaded the dataset, splatted it into X and Y, and fit these parameters in the decision tree regressor to predict prices. We have stored the price with their respective crops in the crop list array to show results on the web page. Accuracy could range from 90 to 92% in terms of price prediction. We have predicted the price for the future one year for each crop in our dataset.

5. Future Scope

These two graphs show India's future and past 1 year

of price prediction for wheat.



Previous 12-month price prediction of wheat

6. Conclusion and Future

The outcomes explain that we can achieve accurate crop price forecasting using the decision tree regressor. The decision tree algorithm predicted the accurate price of the crops with the lowest cost in terms of the performance of crop yield models with the lowest models. It is helpful for significant crop price prediction in farming outlining. It helps farmers make the best decision for the correct crop selection with a higher price in the market. In this paper, we explain the crop price prediction capability of the decision tree algorithm. In the future, we can add more crops feature like soil condition, PH, humidity, geo-location and temperature needed to cultivate a particular crop to predict the more efficient result with much information.

References

[1]. Ahamed, A.T.M.S., Mahmood, N.T., Hossain, N., Kabir, M.T., Das, K., Rahman, F., Rahman, R.M., 2015. Applying data mining techniques to predict the annual yield of major crops and recommend planting different crops in different districts in Bangladesh. In: 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2015 - Proceedings, https://doi.org/10.1109/SNPD.2015.7176185.

[2]. Ahmad, I., Saeed, U., Fahad, M., Ullah, A., Habib- ur-Rahman, M., Ahmad, A., Judge, J., 2018. Yield forecasting of spring maize using remote sensing and crop modelling in Faisalabad-Punjab Pakistan. J. Indian Soc. Remote Sens. 46 (10), 1701–1711. https://doi.org/10.1007/s12524-018-0825-8.

[3]. Cakir, Y., Kirci, M., Gunes, E.O., 2014. Yield prediction of wheat in the southeast region of Turkey by using artificial neural networks. In: 2014 The 3rd International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2014. https://doi.org/10.1109/Agro-Geoinformatics.2014.6910609.

[4]. Gandhi, N., Armstrong, L.J., 2016b. A review of the application of data mining techniques for decision making in agriculture. Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics, https://doi.org/10.1109/ IC3I.2016.7917925.

[5]. Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., Shearer, S., 2018. Integration of high resolution, remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. Comput. Electron. Agric. 153, 213–225. https://doi.org/10.1016/J.COMPAG.2018.07.016.

[6]. Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: a review. Sensors (Switzerland) 18 (8). https://doi.org/10.3390/ s18082674

[7]. Matsumura, K., Gaitan, C.F., Sugimoto, K., Cannon, A.J., Hsieh, W.W., 2015. Maize yield forecasting by linear regression and artificial neural networks in Jilin, China. J. Agric. Sci. 153 (3), 399–410. https://doi.org/10.1017/S0021859614000392.

[8]. McQueen, R.J., Garner, S.R., Nevill-Manning, C.G., Witten, I.H., 1995. Applying machine learning to agricultural data. Comput. Electron. Agric. 12 (4), 275–293. https://doi/10.1016/0168-1699(95)98601-9.

[9]. Osman, T., Psyche, S.S., Kamal, M.R., Tamanna, F., Haque, F., Rahman, R.M., 2017. Predicting early crop production by analyzing primary environment factors, pp. 470–479. https://doi.org/10.1007/978-3-319- 49073-1_51.

[10]. Ranjan, A.K., Parida, B.R., 2019. Paddy acreage mapping and yield prediction using sentinel-based optical and SAR data in Sahib Ganj district, Jharkhand (India). Spatial Inf. Res. https://doi.org/10.1007/s41324-019- 00246-4.

[11]. Shah, A., Dubey, A., Hemnani, V., Gala, D., Kalbande, D.R., 2018. Smart Farming System: Crop Yield Prediction Using Regression Techniques. Springer, Singapore, pp. 49–56. https://doi.org/10.1007/978-981-10-8339-6˙6.

[12]. Taherei-Ghazvinei, P., Hassanpour-Darvishi, H., Mosavi, A., Yusof, K.W., Alizamir, M., Shamshirband, S., Chau, K., 2018. Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. Eng. Appl. Comput. Fluid Mech. 12 (1), 738–749. https://doi.org/10.1080/

[13]. You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep Gaussian process for crop yield prediction based on remote sensing data. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), 4559–4566. https://doi.org/10. 1109/MWSCAS.2006.381794.

[14]. Wang, A., Tran, C., Desai, N., Lobell, D., n.d. Deep transfer learning for crop yield prediction with remote sensing data. Dl.Acm.Org. Retrieved from https://dl.acm.org/ citation.cfm?id=3212707