Crop-yield and Price Forecasting using Machine Learning

2 authors:

Sadiq A Mulla
Secab institute of engineering and technology Karnataka India

1 PUBLICATION   **0** CITATIONS

SEE PROFILE

S. A. Quadri
Universiti Sains Malaysia

32 PUBLICATIONS   **142** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Multisensor Data Fusion View project

Crop-yield and Price Forecasting using Machine Learning

[1]Sadiq A Mulla&[2]Dr.S.A.Quadri

[1]M-Tech, Department of Computer Network Engineering, SECAB Engineering and Technology,

Vijaypur.SECAB Institute Of Engineering and Technology, Karnataka, India,

Email id :-sadiqmulla8@gmail.com

[2]Assistant Professor, Department of Computer Network Engineering, SECAB Engineering and Technology,

Vijaypur.SECAB Institute Of Engineering and Technology, Karnataka, India,

Email id :-reachquadri@yahoo.com

Abstract

Guaranteeing food profitability is a significant issue for the creating nations like India, where more than 33% of the individuals is live in neediness. Season crop yield estimations are inconceivably recognized as a significant contribution for exploring food accounting reports and yield deficiencies. Harvest yield estimation and evaluation is done worldwide on a territorial premise to empower high yield and cost estimation. To estimate cost there is no system in place to advice farmers what crops to grow. Hence, In this paper we attempt to predict crop price that a farmer can obtain from his land, by analysing patterns in past data. We have considered few rabi and kharif season crops like paddy, arhar, bajra, barley etc for our analysis. We make use of several data such as rainfall, temperature, market prices, area of land and past yield of a crop.

In this project, we implement a supervised machine learning algorithm namely, Decision tree algorithm and analyse the data and predict for the new set of data. We also predict the price and the gain for next twelve months over the past twelve months and give the time series analysis of the same.

Keywords*Crop-yield, Supervised Machine learning in Agriculture,Decision tree, Forecasting,*

I. Introduction

Agriculture is that the most vital sector of the Indian Economy. Indianagriculture sector accounts for 18% of India's gross domestic product and provides employment to five hundredth of the country's hands. however latest studies have shown a gentle decline within the contribution created by agriculture to the Asian nation economy though it's demographically the broadest economic sector and plays a major role within the overall socio-economic cloth of India.

An attainable reason for the poor contribution of the agricultural sector to the gross domestic product of Asian nation is also the dearth of adequate crop designing by farmers further as by the govt. The gross domestic product is one in all the most indicators that wants to notice the health of a country's economy. fast fluctuations in crop costs area unit common within the market. These fluctuation in costs is especially owing to lack of previous designing. This ends up in fluctuation within the production and even the value of a crop within the market. It will lead the crops to be extremely priced, being a drawback for the patron, once the value hikes and farmers to suffer with lose in investment once the value drops. In such a state of affairs, it's tough for a farmer to create an informed alternative of crop to grow in his land or to estimate the yield and worth to expect from it. The intention of this project is to assist the farmer build higher selections, by analysing historical yield and worth knowledge victimisation machine learning.

Machine Learning is an application of Artificial Intelligence that has proven to produce good prediction models in various aspects such as stock market, weather, outcome of decisions, crop, and in our case crop yield and price. Machine learning algorithms are divided into three main groups [1] based on their purpose:
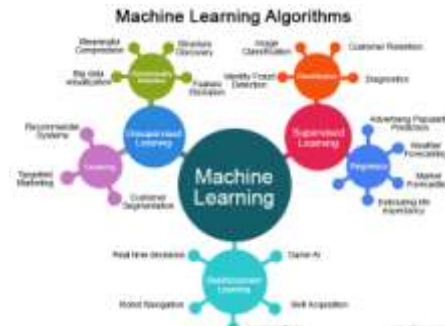
Figure 1.1 Type of Machine Learning

Today the farmers cultivate crops supported the expertise gained from the previous generation. Since the normal technique of farming is practiced there exists Associate in Nursing excess or deficiency of crops while not meeting the particular demand. The farmers aren't aware of the demand that takes place within the current agricultural economy. This ends up in the loss to the farmers. The expressed reasons so as of importance behind farmer suicides were – atmosphere, low turnout costs, stress and family responsibilities, poor irrigation, and increase within the price of cultivation. the most reason is that the low costs of the merchandise and therefore the accumulated price of cultivation. the value of crops is decided by economic demand and therefore the limits of the assembly.

Yield prediction is a vital agricultural drawback. each farmer is curious about knowing, what proportion yield he's concerning expect. within the past, yield prediction was performed by considering farmer's previous expertise on a specific crop. the amount of knowledge is big in Indian agriculture. info} once become information is very helpful for several functions.

In this paper the main aim is to create a user-friendly interface for farmers, which gives the analysis of several crop production based on available data. Different Machine learning techniques were used to predict the crop yield for maximizing the crop productivity.

II. Background

Nowadays, many researchers have implemented several models to predict crop yield predictions. But these models have several drawbacks due to the incorrect use of the algorithms.

In [2], a predictive model is proposed to identify the crop yield using a Hybrid neural network considering the soil parameters and the external climate conditions. But the model fails to predict for the real time fluctuations in the climate and soil parameters. Hence the model does suit for real-time analysis.

In [3], A Prediction system is developed using KNN and Apriori algorithm to analyse and recommend the crops to the farmers. The system has a well-developed Interface to input crop, which allows the user to input the crop names and which it outputs the crop yield. But the system does not have the provision to prediction price at the same time.

In [4], An automated farming crops prediction system is developed using KNN algorithm and Multi-linear regression for Bangladesh countries. But the model is considered to be an initial step in the advancement since it doesn't identify any new research gaps.

In [5], The system is designed to predict crop yield and the fertilizer recommendation which could be used for analysis soil and the current growth of crops, based on which the fertilizer is recommended which is not suitable for the real time analysis.

III. Objectives

The central focus of this research work is to predict crop price based on the previous trends in weather, yield and price. In agri-based country like India, predicting crop yield and price can aid a lot of people whose sole survival is dependent on the crop that they plan to grow.
The major objectives of the proposed system include

- To provide the farmer with the yield of a crop, based on rainfall and WPI using machine learning.
- To predict the future market price of crops by taking previous crop price and sowing month into consideration.
- To predict the crop price for all around the country for next 1 year.

IV. Proposed Work

This work focuses in investigating the prediction of Crop yield and the cost estimation. The proposed methodology uses the tree algorithm in order to predict the results efficiently and proves to best suitable for the research work. The data collected, is analysed and worked to predict the yield and the cost of the crops at any given time.

The research work majorly involves the following implementation modules.
1. Data Acquisition
2. Data Exploration
3. Machine learning prediction
4. Web application

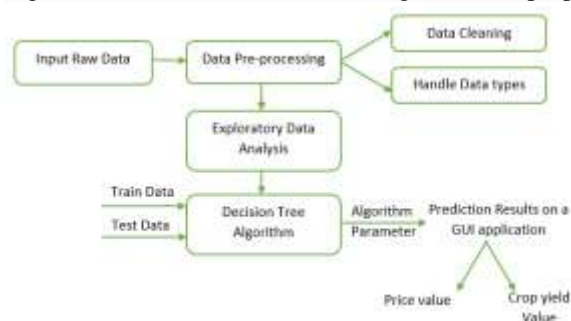Figure 4.1 shows the architecture diagram of the proposed work.

Figure 4.1 Proposed System

Module 1: Data Acquisition

Dataset is prepared by collecting the crop data obtained from the public repository. There are a handful of datasets which contain data. We obtained the data which contains the details of the rainfall of the individual crops.

A sample of acquired set of data and their attribute are shown in the figure 4.2 below, where WPI is represented as Wholesale Price Index.

Month	Year	Rainfall	WPI
4	2012	47.5	104.2
5	2012	31.7	104.2
6	2012	117.8	107
7	2012	250.2	111
8	2012	262.4	114
9	2012	193.5	116.1
10	2012	58.7	116.6
11	2012	30.7	116.3
12	2012	11.7	115.9
1	2013	11.3	116.7
2	2013	40.1	117.9
3	2013	15.7	118.7
4	2013	30.4	120.4
5	2013	57.8	121.9
6	2013	219.8	124.9
7	2013	310	129.4
8	2013	254.7	132
9	2013	152.7	131.3
10	2013	129.4	131
11	2013	14	131.3

Figure 4.2 Dataset

Module 2: Data Exploration

Exploratory Data analysis (EDA) is an essential advance that happens when component coming up with and getting info and it need to be done before any demonstrating. this can be on the grounds that it's

essential for Associate in Nursing info research worker actually to virtually comprehend the thought of the data while not creating suspicions. The when impact of knowledge investigation are often terribly useful in obtaining a handle on the structure of the data, the appropriation of the qualities, and therefore the distance of extraordinary qualities and interrelationships within the informational index.

The purpose of EDA is:

- To utilize summary measurements and representations to all ora lot of seemingly comprehend data, discover items of data regarding the inclinations of the data, its quality and to detail suppositions and therefore the speculation of our hypothesis.

- For knowledge preprocessing to be effective, it's basic to own a general image of your data Basic factual portrayals van utilized to differentiate properties of data and |the knowledge and have that information esteems got to be treated as commotion or exceptions. Next step is to explore the information. There square measure 2 approaches want to examine the information using: Descriptive statistics is that the method toward gathering key attributes of the informational index into easy numeric measurements. a little of the regular measurements utilized square measure mean, variance, and relationship.

- Visualization is that the method toward anticipating the data, or components of it, into mathematician area or into dynamic photos. within the data mining method, data investigation is used during a big selection of steps together with preprocessing, modelling, and interpretation of results.

During this process of analysis, Univariate and Bivariate analysis is done. Figure 4.3 shows the distribution of rainfall across all the states of India.
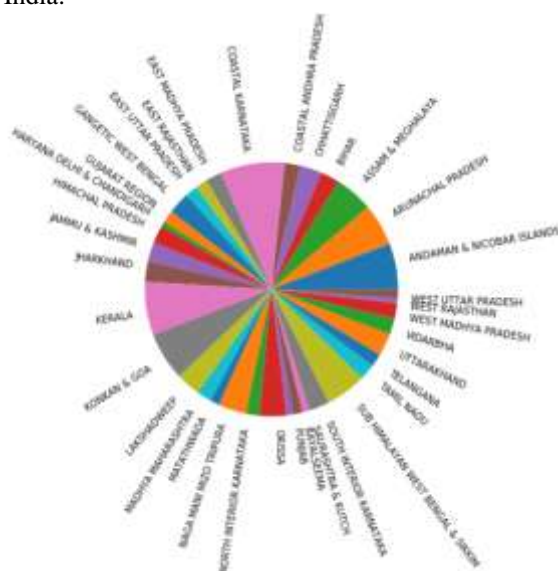

Figure 4.3 Distribution of Rainfall across India

Figure 4.4 represents the Top 3 crops of the Kharif season crops produced in the highest crop producing states.
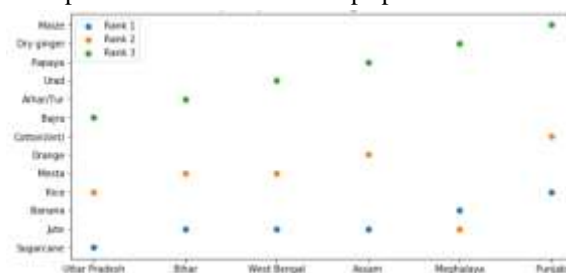

Figure 4.4 Top 3 crops of Kharif season Crop Distribution

Figure 4.5 represents the Top 3 crops of the Rabi season crops produced in the highest crop producing states.
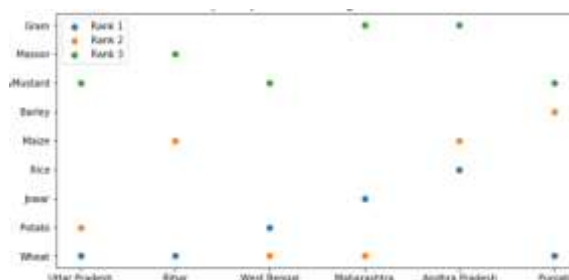
Figure 4.5 Top 3 crops of Rabi season Crop Distribution

Module 3: Machine learning Prediction

Machine learning prediction has these following steps:
1. Divide datainto 2 parts: trainingandtesting data.
2. Defining the algorithms namely Decision tree algorithm.
3. Training and testing against the algorithms.
4. Updating the User Interface with the calculated values.

Decision tree Algorithm

Decision tree is associate degree formula that uses a tree like graph or model   and their potential outcomes to predict the ultimate decision, this formula uses conditional management statement. a call tree is associate degree formula for approaching discrete-valued target functions, during which call tree is denoted by a learned perform. For inductive learning these styles of algorithm's area unit terribly notable and are with success applied to abroad vary of tasks. we tend to offer label to a brand new dealing that's whether or not it's legit or fraud that category label is unknown then dealing worth is tested against the choice tree, and at that time from root node to output/class label for that dealing a path is copied. call rule determines the result of the content of leaf node. normally rules have the shape of 'If condition one and condition two however not condition three then outcome'. call tree helps to work out the worst, best and expected values for various situations, simplified to grasp and interpret and permit addition of latest potential situations. Steps for creating a call tree area unit that foremost to Calculate the entropy of each attribute victimisation the dataset in downside then dataset is split into subsets victimisation the attribute that gain is most or entropy is minimum at that time to form a call tree node containing that attribute and in conclusion rule is performed on subsets victimisation remaining attributes to make a call tree.
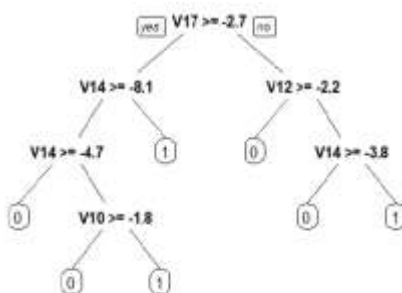

Figure 4.6 Decision tree sample

V.   Results and Evaluation

In the evaluation, we want to understand, for a number of metrics, whether our method works well for the problem statement we are trying tackle. We calculate the crop yield, its increase or decrease and also its price. Figure 5.1 shows the top gaining crops obtained using Decision tree algorithm.

Figure 5.1 Top 5 gaining crops

Figure 5.2 shows the top losing crops obtained using Decision tree algorithm.

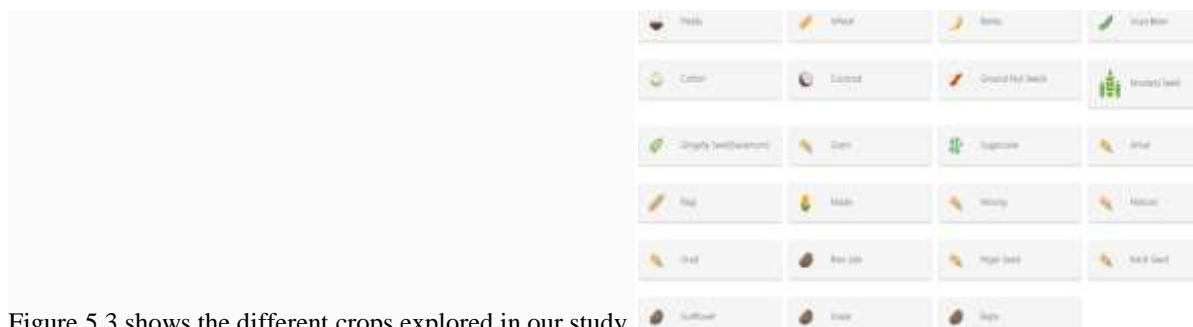Figure 5.2 Top 5 losing crops

Figure 5.3 shows the different crops explored in our study.

Figure 5.3 Kharif/Rabi crops

Figure 5.4 shows the 12month forecast for Rice crop, where we do the same for all the crops.

Figure 5.4 Crop Yield/Price Forecast

Figure 5.5 shows the next and previous 12 month forecast time series graph for Rice crop, where we do the same for all the crops.
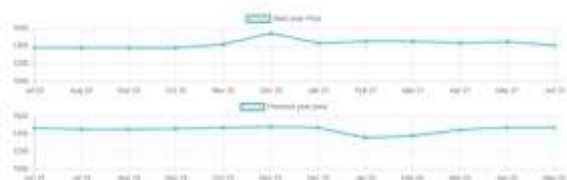

Figure 5.5 Time-series Yield/Price Forecast

VI. Conclusion

The proposed model has been developed to help farmers make better decisions with respect to which crop is most suitable during his desired time of sowing and the location. Our system predicts the yield and price of the crop of choice, giving the farmer useful information well before starting the process of cultivation.

Numerous prediction algorithm can be used for crop yield and price prediction such as decision trees, neural networks, SVM etc. Our model uses Decision tree. It is trained on several kharif and ragi crops (like paddy, arhar, bhajra, barley, etc) providing a good accuracy.

VII.   Future Work

To be able to responsibly feed a growing population, it's vital that farmers increase food production on existing farmland to avoid deforestation. the longer term work done ought to optimize farming practices to extend yields, crop quality and incomes in an exceedingly property manner.

An Advanced value prognostication system may be developed wherever a dashboard can predict the market value trends exploitation statistical method for a minimum of a period and therefore the production pattern of various crops. A Platform for Agriculture ought to use massive information, AI, Machine Learning, satellite representational process and weather information to assess the land area and monitor crop health on a true time basis. so it will notice cuss and sickness infestations, estimate the crop output and yield, and conjointly forecast costs.

Other key input such as the prices in major markets of neighbouring States can also be factored into the price forecast. The System should be developed in a way that it does a real time analysis to get an accurate solution to farmers problems and optimize their farming practices. The system can introduce and make available climate-aware cognitive farming techniques and identifying systems of crop monitoring, early warning on pest/disease outbreak based on advanced AI innovation.

References

[1] https://www.educba.com/machine-learning-algorithms/

[2] S. Kulkarni, S. N. Mandal, G. S. Sharma, M. R. Mundada and Meeradevi, "Predictive Analysis to Improve Crop Yield using a Neural Network Model," *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Bangalore, 2018, pp. 74-79.

[3] S. V. Bhosale, R. A. Thombare, P. G. Dhemey and A. N. Chaudhari, "Crop Yield Prediction Using Data Analytics and Hybrid Approach," *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, 2018, pp. 1-5.

[4] T. Siddique, D. Barua, Z. Ferdous and A. Chakrabarty, "Automated farming prediction," *2017 Intelligent Systems Conference (IntelliSys)*, London, 2017, pp. 757-763.

[5] S. Bhanumathi, M. Vineeth and N. Rohit, "Crop Yield Prediction and Efficient use of Fertilizers," *2019 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, 2019, pp. 0769-0773.

[6] P. Bose, N. K. Kasabov, L. Bruzzone and R. N. Hartono, "Spiking Neural Networks for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6563-6573, Nov. 2016.

[7] N. Hemageetha and G. M. Nasira, "Radial Basis Function Model for Vegetable Price Prediction", International Conference on Pattern Recognition, Informatics and Mobile Engineering, pp. 424-428, 2013

[8] Yung-Hsing Peng, Chin-shun Hsu and Po-Chuang Huang, "An investigation of Spatial approaches for crop price forecasting in different Taiwan markets",

[9] Prof A K Mariappah and J Austin Ben Das, "A Paradigm for rice yield prediction in Tamil Nadu", IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development (TIAR), pp. 18-21, 2017

[10] J. Hartigan, Clustering Algorithms, John Wiles & Sons, New York, 1975.

[11] A. Mucherino, A. Urtubia, Consistent Bi clustering and Applications to Agriculture, IbaI Conference Proceedings, Proceedings of the Industrial Conference on Data Mining (ICDM10), Workshop "Data Mining in Agriculture" (DMA10), Berlin, Germany, 105-113, 2010.

[12] Fagerlund S Bird species recognition using Support Vector Machines. EURASIP J Adv Signal Processing, Article ID 38637, p 8, 2007.

[13] Holmgren P, Thuresson T Satellite remote sensing for forestry planning: a review. Scand J For Res 13(1):90– 110, 1998.

[14] Das KC, Evans MD Detecting fertility of hatching eggs using machine vision II: Neural Network classifiers. Trans ASAE 35(6):2035–2041, 1992.