PAPER • OPEN ACCESS

Crop price prediction using supervised machine learning algorithms

To cite this article: Ranjani Dhanapal *et al* 2021 *J. Phys.: Conf. Ser.* **1916** 012042

View the article online for updates and enhancements.

Retraction

Retraction: Crop price prediction using supervised machine learning algorithms (*J. Phys.: Conf. Ser.* **1916** 012042)

Published 23 February 2022

This article (and all articles in the proceedings volume relating to the same conference) has been retracted by IOP Publishing following an extensive investigation in line with the COPE guidelines. This investigation has uncovered evidence of systematic manipulation of the publication process and considerable citation manipulation.

IOP Publishing respectfully requests that readers consider all work within this volume potentially unreliable, as the volume has not been through a credible peer review process.

IOP Publishing regrets that our usual quality checks did not identify these issues before publication, and have since put additional measures in place to try to prevent these issues from reoccurring. IOP Publishing wishes to credit anonymous whistleblowers and the Problematic Paper Screener [1] for bringing some of the above issues to our attention, prompting us to investigate further.

[1] Cabanac G, Labbé C and Magazinov A 2021 arXiv:2107.06751v1

Retraction published: 23 February 2022

Crop price prediction using supervised machine learning algorithms

Ranjani Dhanapal [1], A AjanRaj [1], S Balavinayagapragathish [1], J Balaji [1]

[1] Department of Information Technology, Sri Krishna College of Technology, Coimbatore, India

d.ranjani@skct.edu.in

Abstract. Our target is focused largely on agriculture. In agriculture, farmers play the most important role. When the price falls after the harvest, farmers face immense losses. A country's GDP is affected by the price fluctuations of agricultural products. Crop price estimation and evaluation are done to take an intelligent decision before farming a specific type of crop. Predicting the price of a crop will help in taking better decisions which results in minimizing the loss and managing the risk of price fluctuations. In this paper, we predicted the price of different crops by analyzing the previous rainfall and WPI data. We used the decision tree regressor (Supervised machine learning algorithm) to analyze the previous data and predict the price for the latest data and estimate the price for the twelve months to come.

Keywords: price prediction, decision tree, crop price, regression, forecasting, machine learning

1. Introduction

In our country, agriculture is the principal pillar of the economy. The majority of families are dependent on agriculture. The country's GDP is primarily focused on agriculture. More than half of the land is used for agriculture to meet the needs of the population of the region. It is necessary to modernize agricultural practices to meet the demanding requirements. Our research aims to solve the problem of crop price prediction more effectively to ensure farmers' incomes. To come up with better solutions, it uses Machine Learning methods on different data.

Productivity can be improved by understanding and predicting crop prices through this application. An efficient crop price forecasting system can offer farmers opportunities that can benefit people in a larger context. The fast fluctuations in crop costs are common within the market. These fluctuations in costs are especially owing to the lack of previous design. This leads to fluctuations in demand and also in the market value of a crop. Once the value rises and farmers suffer from an investment loss after the value decreases, it will lead the crops to be highly-priced, becoming a disadvantage for consumers. Farmers are not aware of the demand within the emerging agricultural economy that is taking place. Machine learning can be defined as one of the Artificial Intelligence applications that have proven to produce successful prediction models in various aspects, such as the stock market, weather, business decisions, and crop prices in our case. Eventually, the findings are displayed as a web application so that farmers can easily access them.

Farmers are not any longer looking to use analytics to get data they need to realize actionable insights and take intelligent decisions. Most of the farmers in other countries are started to migrate for automated farming. The Decision Tree algorithm belongs to the family of learning algorithms that are supervised. The purpose of using a decision tree is to build a training model that will be used by training basic decision

rules derived from previous data to predict the category or meaning of the target variable. Tahmid Shakoor's paper helped us study various attributes like soil capability, depth of soil, humidity, rainfall, soil texture, topsoil erosion, and permeability [1]. A model was featured for crop prediction, it executes well for only certain data the fluctuations in the real world make the model a failure one. We use a decision tree regressor with an exact dataset to provide the expected results. The other advantage is that the system has an annual rainfall forecast feature, which in this case originates from WSP data. The ability to predict crops before the start of the crop season. This allows users the opportunity to make policy adjustments to accommodate drastic market fluctuations further forward in the crop cycle, such as selecting lucrative crops before planting or even changing the type of crop [2].

Another model uses the KNN algorithm to predict the crop price, the model had a got UI and users can easily search for crops and it also uses climate data but the model is unable to predict for long period. The algorithm we developed can predict the price of the crops for 12 months by which the farmers can know the future price of the crops and plant according to it [3]. Although several methods are available to obtain predictions of rainfall, the algorithm described in this paper succeeded in emphasizing rainfall along with the prediction of crop yield [4]. A model is developed for predicting the price of a fertilizer they used crop prediction to analyze the fertilizer price, but the fertilizer provided is not recommended as it contains many poisonous chemicals that can harm consumers [5]. The predicted results are more beneficial to the food chain of the country since we can predict the price even before sowing the seeds in the ground [6-8]. This model can also be used by the government, panchayat unions, financial investors so that they can plan the money they need to spend beforehand. In [9], The K Nearest Neighbors algorithm and Multi-linear regression for Bangladesh are used to create an automatic farming crop prediction method [10]. However, because it does not recognize any new research differences, the model is considered an initial phase in advancement [11]. The model we identified is scalable as it can provide a prediction for more than 20 types of crops. The price, when compared to last year, can be easily found via the graph, which is simple and self-explanatory. The model is helpful to all kinds of farmers, independent of the size of their farming land. We are improving the model to predict the price of each crop in the world and give fertilizer suggestions to it so that the farmers can yield more crops than expected.

2. Decision tree regressor based price predictor

This work focuses on investigating the prediction of Crop price estimation. The proposed methodology uses the decision tree algorithm to predict the results efficiently and proves to best suitable for the research work. The data collected, is analyzed and cleaned to predict the price of the crops. The architecture of the proposed crop prediction system is depicted below in Figure 1.
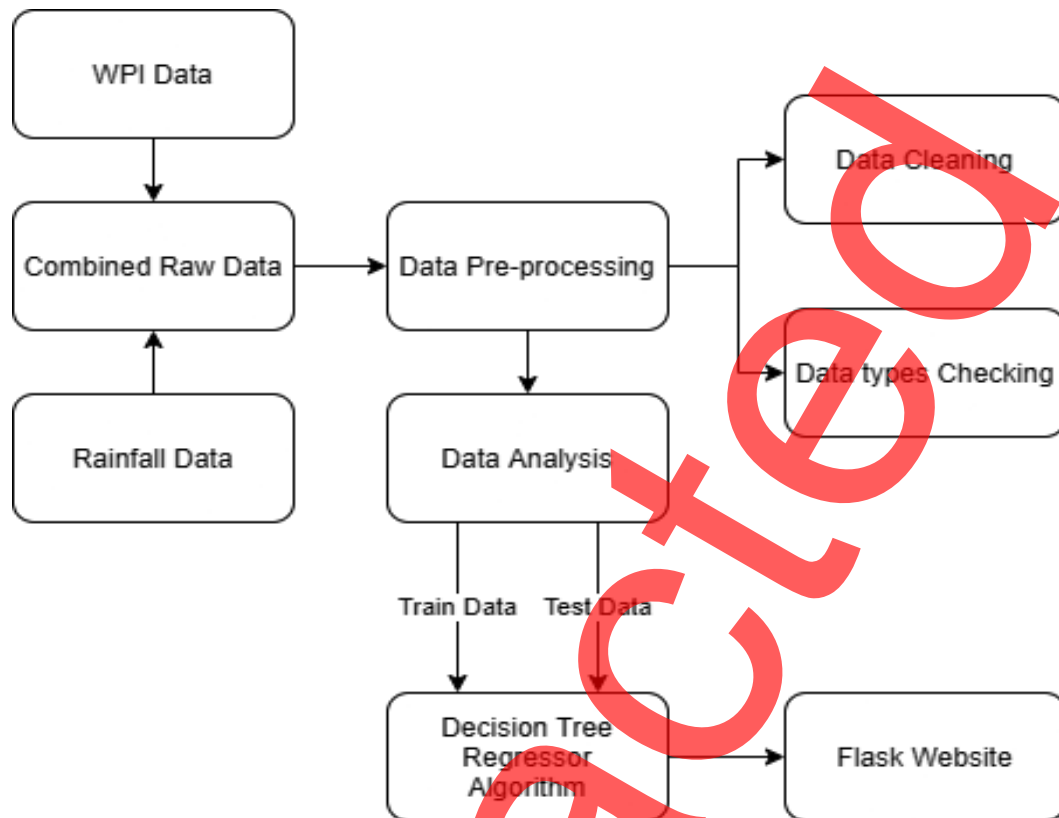
Figure 1. Crop price prediction system architecture

The implementation is divided into the following modules.

    A. Data Gathering

    B. Data Cleaning

    C. Data Exploration

    D. Prediction using Machine learning

    E. Web application

1. Data Gathering

Dataset is prepared by collecting the crop and rainfall data from the Indian government data repository (data.gov.in). There are a lot of datasets that contain data. We obtained the data which contains the details of the whole price index and rainfall of the individual crops per month Figure 2.

Barley			
Month	Year	Rainfall	WPI
4	2012	47.5	117.1
5	2012	31.7	118
6	2012	117.8	111.2
7	2012	250.2	113.2
8	2012	262.4	113.8
9	2012	193.5	111.8
10	2012	58.7	110.5
11	2012	30.7	113.9
12	2012	11.7	119
1	2013	11.3	119.4
2	2013	40.1	120.2
3	2013	15.7	119.5
4	2013	30.4	116.5
5	2013	57.8	114.7
6	2013	219.8	115.8
7	2013	310	115.8
8	2013	254.7	116.3
9	2013	152.7	118
10	2013	129.4	118.6
11	2013	14	120.1
12	2013	6.7	123.3

Figure 2. Combined dataset of Barley

2. Data Cleaning

One of the most significant steps in any machine learning project is data cleaning. There are several different methods of statistical analysis and data visualization techniques in the dataset that you can use to explore the data to identify the appropriate data cleaning operations to be conducted. There are some very simple data cleaning operations before jumping to the advanced methods that we can conduct in a machine learning project on every single dataset. They are so important that models can break or report excessively optimistic outcomes of success if missed. In our dataset, we cleaned all the null values and checked whether all the datatypes are valid.

3. Data Exploration

Also known as E.D.A, exploratory data analysis is a very important phase in researching and investigating various data sets and summarizing their significant characteristics, often using different methods of data visualization. It allows it simpler for a data analyst to obtain repeated trends, spot anomalies, test theories, and conclusions to decide the best way to monitor data sources to get the results with greater precision.

4. Prediction using Machine learning

We tested around 6 different algorithms and found the Decision Tree Regressor algorithm as the most effective.

Its root mean squared value is 3.8 which is very less when compared to other algorithms

DecisionTreeRegressor

```
clf = DecisionTreeRegressor()
clf.fit(X_train, y_train)
print("DecisionTreeRegressor")
y_pred = clf.predict(X_test)
df = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
df
print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
```

```
DecisionTreeRegressor
Mean Absolute Error: 2.4294117647058804
Mean Squared Error: 15.142941176470595
Root Mean Squared Error: 3.8913932179195916
```

Figure 3. Root mean square of Decision tree regressor

A Decision Tree is one of the most commonly used algorithms for supervised learning. In the form of a tree structure, a decision tree generally generates regression models or classification models. It breaks down a dataset into smaller subsets and, based on the subsets, a decision tree is created. A tree containing decision nodes and leaf nodes is the final product. There are mostly two or more branches of a decision node, each representing values for the checked attribute. An option for the final numerical goal is represented by the Leaf node. The topmost node in the tree corresponds to the root node, which is the best node. Both continuous and categorical data can be processed by decision trees Figure 3.

Machine learning prediction has the following steps:

Step 1: Initialize a dataset containing information on rainfall and wholesale price index.

Step 2: From the dataset select all the rows and columns 1,2,3 to "X" Which is the independent variable.

Step 3: From the dataset select all of the rows and column 4 to "y" Which is the dependent variable.

Step 4: Fit the x and y variables with a decision tree regressor.

Step 6: Update the UI with predicted values.

5. Web application

The Predicted WPI data is converted into Price per Quintal using the formula (WPI x Base Price)/100 and displayed in a visually understandable web application created using the Flask framework. Flask is one of the popular, extensible web micro-framework for building beautiful web applications with Python. An Index page is created and from there we can navigate to 20 different crops and see their forecast in detail Figure 4.
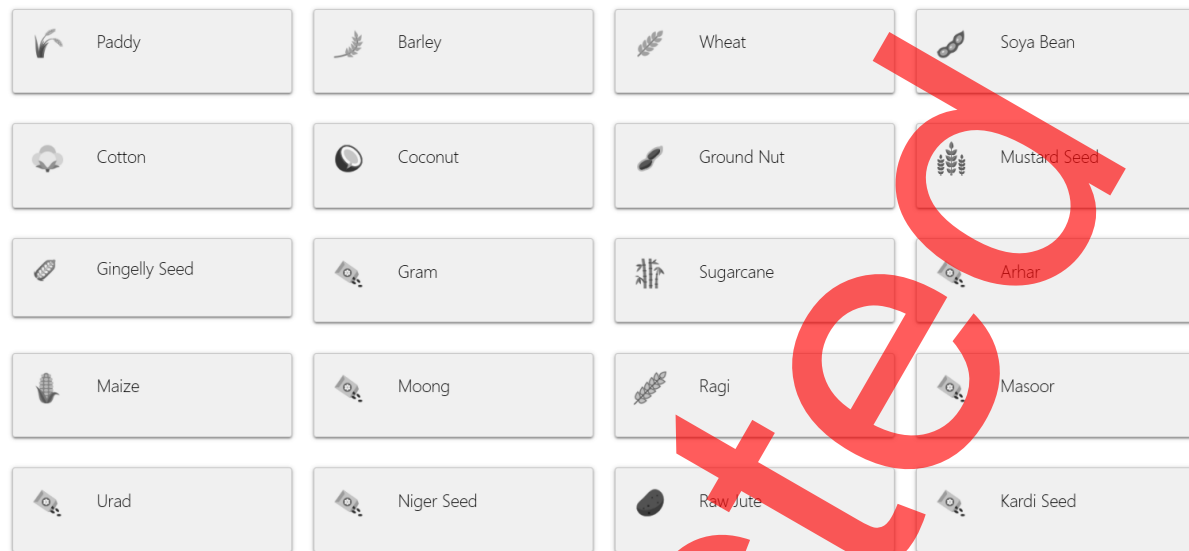
Explore by Crops

Paddy	Barley	Wheat	Soya Bean
Cotton	Coconut	Ground Nut	Mustard Seed
Gingelly Seed	Gram	Sugarcane	Arhar
Maize	Moong	Ragi	Masoor
Urad	Niger Seed	Raw Jute	Kardi Seed

Figure 4. Index Page

3.Implementation and result:

In the evaluation, we need to understand, several metrics, whether our decision tree model works well for the problem statement. We calculate the crop price, its increase, and decrease percentage, and also its 12-month forecast Figure 5-8.

Top Gainers

Item Name	Price (per Qtl.)	Change
Soyabean	₹3348.4	11.42% ▲
Niger	₹6657.0	6.02% ▲
Ragi	₹3223.5	3.37% ▲
Copra	₹10934.4	2.39% ▲
Arhar	₹3728.0	1.39% ▲

Figure 5. Top Gaining crops

Top Losers

Item Name	Price (per Qtl.)	Change
Jute	₹2793.9	-9.74% ▼
Barley	₹1381.8	-8.14% ▼
Sunflower	₹3722.2	-3.64% ▼
Moong	₹3955.0	-3.42% ▼
Bajra	₹1502.82	-2.89% ▼

Figure 6. Top Losing crops

12 Month Forecast

Month	Price (per Qtl.)	Change
Mar 21	₹1129.94	-1.03% ▼
Apr 21	₹1142.68	0.09% ▲
May 21	₹1135.82	-0.52% ▼
Jun 21	₹1109.36	-2.83% ▼
Jul 21	₹1089.76	-4.55% ▼
Aug 21	₹1085.84	-4.89% ▼
Sep 21	₹1085.84	-4.89% ▼
Oct 21	₹1072.12	-6.09% ▼
Nov 21	₹1117.2	-2.15% ▼
Dec 21	₹1214.22	6.35% ▲
Jan 22	₹1126.02	-1.37% ▼
Feb 22	₹1141.7	0.0% ▲

Figure 7. Price Forecast of Barley

Figure 8. Time-series Price Forecast

4. Conclusion and future enhancement

In this paper predicting the price of crops and forecast the price of the next 12 months is proposed. The data is presented via a flask web page and runs on effective machine learning algorithms and technologies with an overall user-friendly interface for users. The acquired training datasets provide ample insights to forecast the required price and demand in the markets. Therefore, the scheme allows farmers to reduce their problems and raise their income.

Various algorithms can be used for crop price prediction such as decision trees, support vector machines, neural networks, deep learning, etc. Our model used a supervised machine learning algorithm called Decision tree Regressor. It is trained on several Kharif and Ragi crops (Paddy, Wheat, Cotton, Barley, etc.) providing better accuracy.

The Model further can be trained with climate-aware farming techniques, provide fertilizer suggestions, and identifying systems of crop monitoring, warning on pest outbreak, disease outbreak based on advanced AI Models.

References

[1] S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh.2019. Machine learning approach for forecasting crop yield based on climatic parameters .978-1-4799-2352- 6/14/$31.00 ©2014 IEEE

[2] Takeshi Yoshida Noriyuki Murakami and Hiroyuki Tauiji.2017. Hybrid Machine Learning Approach to Automatic Plant Phenotyping in Smart Agriculture. 978-1- 5090-5888-4/16/$31.00 @IEEE 2016.

[3] Neha Rale, Raxitkumar Solanki, Doina Bein, James Andro-Vasko, Wolfgang Bein. Prediction of Crop Cultivation. 978-1-7281-0554-3/19/$31.00©2019 IEEE.

[4] Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty.2017. Agricultural Production Output Prediction Using Supervised Machine Learning Techniques.978-1-5386-3831-6/17/$31.00 ©2017 IEEE.

[5] G Srivatsa Sharma, Shah Nawaz Mandal, Shruti Kulkarni, Monica R Mundada, Meeradevi.2018. Predictive Analysis to Improve Crop Yield Using a Neural Network Model. 978-1- 5386-5314-2/18/$31.00 ©2018 IEEE.

[6] Talha Siddique, Dipro Barus, Zanntual Fredous, Amitabh Chakravarti. 2017. Automated Farming Prediction.0978-1- 5090-6182-2/17 @2017 IEEE.

[7] Haldorai, A. Ramu, and S. Murugan, Social Aware Cognitive Radio Networks, Social Network Analytics for Contemporary Business Organizations, pp. 188–202. doi:10.4018/978-1-5225-5097-6.ch010

[8] R. Arulmurugan and H. Anandakumar, Region-based seed point cell segmentation and detection for biomedical image analysis, International Journal of Biomedical Engineering and Technology, vol. **27**, no. 4, p. 273, 2018.

[9] Yung-HsinPeng, Chin-Shun Hsu, and Po-Chuang Huang Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets, 2017 IEEE

[10] Md.TahmidShakoor, Karishma Rahman, SumiyaNasrinRayta, AmitabhaChakrabarty Agricultural Production Output Prediction Using Supervised Machine Learning Techniques, , 2015 IEEE.

[11] T. Siddique, D. Bara, Z. Ferdous and A. Chakrabarty, Automated farming prediction, 2017 Intelligent Systems Conference (IntelliSys), London, 2017, pp. 757-7