

An intelligent Crop Price Prediction using suitable Machine Learning Algorithm

Ishita Ghutake¹, Ritesh Verma¹, Rohit Chaudhari¹, and Vidhate Amarsinh¹

¹Department of Computer Engineering, Ramrao Adik Institute of Technology, Nerul, Navi Mumbai, India

Abstract. Planning of crops for the next season has been a tedious task for the farmers as it is a difficult prediction about metrics of prices that their crop will fetch in a particular season which will be typically based on dynamic weather conditions. This leads to inaccurate prediction of crops' prices by farmers, and they happen to wrongly select the crops or in haste they happen to sell their crops early without storing and thus earning less than what the same crop would have fetched them in the future. This problem could be addressed by an ML model which will predict the prices of crops in advance showing the proper analysis of the crop and presenting their future scenario so that farmers can select the right crops to strategize crop production which involves crop selection, time of sowing deciding crop pattern and storage of harvested crops providing enough insights for predicting the appropriate price in the markets.

Keywords — Prediction, Random Forest Regression, Decision Tree Regression.

1. Introduction

India is known as an Agrarian country as about 55% of India's population depends on agriculture or related activities for their livelihood. Agriculture has a significant contribution to the primary sector of India's economy. At present farmers are facing massive loss, its reasons are infertile soil, climate change, oversupply in the market and there are a lot of other uncertainties involved. Prediction of expected price in the future is very much needed to manage and sell the products at the right time to maximize revenue and minimize loss. This prediction is based on data obtained from government sources and it uses Machine Learning techniques. This website can also potentially benefit other agriculture-based industries to strategize sourcing of raw material at a reasonable price. The research is based on machine learning which will provide farmers and people in allied profession with a prediction which they can access anytime and will present them with predicted price trends in various crops 12 months (monthly basis) in advance. Thus a ML model with good prediction accuracy can prove to be a great asset for our feeding farmers.

2. Motivation and Scope

Our economy is largely based on agriculture and allied professions (about 55% working professionals are employed in this) but still we are not very advance when it comes to analysis based on conditions and produce and smart techniques to have higher yield. The genesis of the research is the idea of crop planning which is very essential more than ever before to have a sustainable climate resilient food system. Crop production is highly subjective specifically the climate conditions. As in this we are considering rainfall as one of the main factors, on the basis of previous data of rainfall and the respective month's prices it would help us produce the accurate price in

upcoming months on the basis of rainfall and correlated price. It's the high time we use technology we harness the potential at algorithms and help to mass increase their profit by selecting the most profitable crop and plan, sowing, harvesting and storage of the crops taking in consideration the predicted price monthly, so that they can select the crops in rotation and plan the storage of harvest so when the respective crops fetches the highest it can be sold.

This research will help farmers to strategize crop production which involves crop selection, time of sowing deciding crop pattern and storage of harvested crops. This also aims at helping people in agriculture allied professions in strategizing their raw material procurement and it is aimed at being very intuitive to the users.

3. Literature Survey

Several approaches have been used in order to improve financial output of agricultural produce. Some noteworthy systems were analyzed and considered in our system development. [1] Data to be used in this system is collected of concerned crops from local markets, online surveys using this collected data as dataset Machine Learning Models are trained. Price prediction is done using algorithms like Artificial Neural Networks, Partial least square & Autoregressive Integrated Moving Average. According to the results obtained by using above mentioned algorithms with the recent data for short and long duration prediction Partial least square and Artificial Neural Networks give better solutions when compared. [2] This paper was aimed at helping farmers take decisions based on ranking of suitability of a crop to concerned area. Prediction and ranking is done using supervised machine learning techniques such as K nearest neighbour regression algorithm and decision tree learning, [3] This paper provides a brief analysis of crop yield predictions using the Multiple Linear Regression process of the selected region.

It focuses on the agricultural analysis of organic and non-organic farming, timely crop cultivation, profit and loss of data and analysis of local business land in a particular area. It focuses on the organic, inorganic and real estate data sets where agricultural forecasts will be available. [4] This paper has shown how data mining techniques can be used in predicting crop yields according to the input parameters. Crop production is influenced by many agro-climatic input parameters. A System built for predicting the crop yield from the given input of climatic parameters indicate a trend of each crop being predominantly influenced by a particular climatic parameter. [5] This paper discusses various applications for data mining in solving various agricultural problems. It integrates the work of different authors in one place, so it is useful for researchers to get details of the current state of data mining techniques and applications depending on the agricultural field.

It also provides research on a variety of data mining techniques used in agriculture including Artificial Neural Networks, K-nearest neighbor, Decision Tree, Bayesian network, Fuzzy set, Support Vector Machine and K – means.

4. Proposed System Methodology

Its going to be website which could be accessed by farmers so that they can base on their financial conditions, need and feasibility and other metrics choose their desired crop. There will be multiple crops widely grown in the country. The dash will show the best doing crop along with the worst and the percentage by which they are soaring or trailing. Predictions will be till 12 oncoming months.

We are creating the Crop Price Prediction Website for crop forecasting were we take data from government of 20+ crops and represent the data in a structured manner representing the increase and decreasing the prices of crops per month and further showing the crops details like its type, location and export factors for the ease of the farmers to plan and manage their finances and sown/harvesting accordingly and also show data in the form of pie chart and graphs. It has user friendly interface and decision tree regression is used for prediction we do the In-Depth statistical analysis of previous data to create the refined platform for interaction which could be accessed by farmers so that they can on the basis of their prediction for 12 oncoming months.

Firstly from data.gov.in updated dataset will be taken comprising of rainfall and wholesale prices respectively month wise for every crop. After required pre-processing model will be trained and then aptly judged. If found suitable, front end and backend will be designed and the ML model will be deployed at the backend. Requisite updating will be timely done on the dataset model will be redeployed.

Here we are doing supervised learning because we have multiple inputs, an output and we are deriving a correlation between them.

The two options suitable for this are linear regression and decision tree regression because both can predict a range of values(continuous) based on multiple inputs and the one we choose is decision tree regression because here by observation in the given dataset there is no linear relation

between the inputs and output. The algorithm will take inputs: The input parameters (months, year and monthly rainfall) Jupiter notebook for training ML model and Visual Studio to develop frontend and backend.

Data ingestion is to be done with the data collected from various sources. The injected data is to be prepared according to the requirement of the system. The Machine Learning model is to be designed and trained using the prepared data. Evaluation of the model is to be done using standard metrics. If the results are not as per the requirements, retrain the model. When the desired results are achieved deploy the system.

Details of Hardware/Software Requirement:

- 1.Jupyter notebook for training ML model and Visual Studio to develop frontend and back- end.
- 2.Python:
 - Flask: It's a web framework used for backend development for the website and linking to html pages using predefined functions.
 - Pandas: To be used to read the dataset and split it into independent, dependent variable and training and testing set
 - NumPy: To shape the data as an array
 - Scikit learn: To use regression algorithms
 - Matplotlib: To plot the decision tree model for visual analysis
- 3.HTML to define the content of web pages
- 4.CSS to specify the layout of web pages
- 5.Java Script for scripting and programming the behaviour of web pages
- 6.csv files to store dataset.
- 7.Chart.js for flexible graphical presentation.

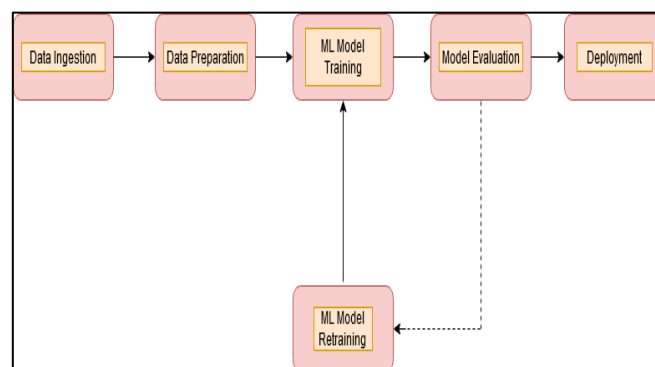


Figure 1: System Design

5. Data Pre-processing

It is a method in which the given raw dataset that is the dataset with unwanted attributes, having missing values, is converted into a processed dataset that is the dataset which is obtained after working on the raw dataset.

Algorithm Selection:

Decision Tree Regression: The dataset will be divided into multiple leaflets which are result of multiple decisions of yes and no and then the new data will be calculated based on what leaflet they land and then calculate the average of that leaflet.

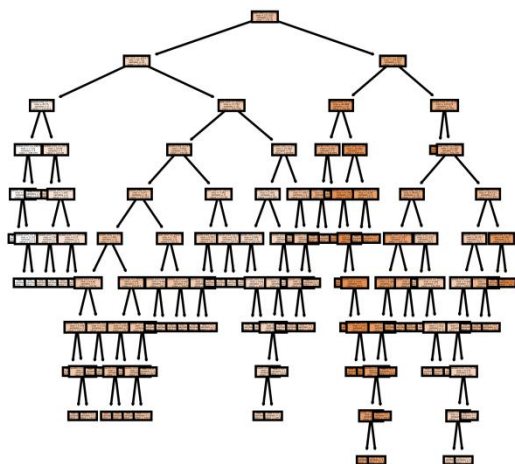


Figure 2: Implementation of Decision Tree Regression

Random Forest Regression:

It works on the basis of ensemble learning, which says that if we combine multiple algorithms or the same algorithm multiple times then we can create a superior algorithm. Random Forest makes use of multiple decision trees to give the output. As we have a huge dataset. Random forest will first extract a small chunk of that data feed it to one decision tree regression model and chain that model this process will be repeated multiple times. We can control by specifying the value of how many decision tree regressions we want. Now we got is a huge dataset on which multiple decision trees are trained. Now the testing data is provided to each of the decision tree, and they will give the output according to themselves and all the collection of output will be then averaged.

Training and Testing:

The dataset is split into two sets of training dataset and testing dataset. 80% for training the dataset & 20% for testing the dataset. The training dataset is a part of data taken for fitting model while the test dataset is also a part of data for the final testing of model on the training dataset. Using the train dataset, we will train our machine learning model. Our machine learning model will attempt to understand and learn on its own and then by using the test dataset we will test our model.

Evaluation:

On evaluating Regression Algorithms such as the Multiple linear, Random Forest and the Decision Tree and testing it on test dataset, made performance analysis of all the models with performance metric and then chosen Decision Tree algorithm.

About Website:

The home page contains various tabs with the names of all the respective crops so that it can be intuitively used by everyone. It's been kept very simplistic so as to make it more user friendly. After clicking on the tabs it will be redirected to the respective page of that crop which contains various details about it.

The second page has multiple sections which describe: Percentage change in price over next 12 months Two graphs indicating the price trends of comparing two consecutive years which are given in the below section.



Figure 3: Homepage



Figure 4: Redirected page

6. Result Analysis

Dataset taken from data.gov.in is analyzed and pre-processed. Two main regression techniques are used to build machine learning prediction model. To show the detailed analysis we consider wheat commodity.

- Current Price: - 1574.1/quintal (As of March 2020)
- Brief forecast (last 6 Months): -
- Min crop price:1476.9/quintal (19 Oct)
- Max crop price: 1672.65/quintal (19 Dec)
- Forecast Trends: We consider the prediction of a single month here to check the accuracy of our model.
- Price of wheat: 1495.8/quintal (Govt data of 19 Sept)
- Price of wheat predicted:1421.45/quintal (Predicted value)
- Change= -4.97% So, for the wheat crop price (19 Sept) we had approximately 95% Accuracy.

And overall, our model possess 92% Accuracy, varying slightly month to month.

7. Conclusion and Future Research

Detailed analysis has been conducted based on real time dataset using two different machine learning techniques. The research aims at a farmer friendly interactive website, predicting the price and forecast through web application and it is runs on efficient machine learning algorithms and technologies having an overall responsive interface to the users. The training datasets so obtained provides enough insights for predicting the appropriate price in the markets. Successfully predicting the price of the crops with 92% accuracy. The future objective is also to interface regional language interface, so that it should become easy to understand and communicate mutually between the software and farmers. The future enhancement also would be to add more features as we are considering rainfall as the factor we can further include temperature, soil fertility and regional use depending upon which the production of the crop varies from area and area so for that a better prediction model can be created and also further enhancement can

done by focusing the accuracy of the model and increasing it to 97% or so.

8. Acknowledgement

Thanks to resources on govt site data.govt.in, Institute Management for providing the resources and helping us in all the possible ways. We also thank readers of this paper for showing interest in this topic and contributing towards the enhancement of this topic as well.

9. References

1. Yung-Hsing Peng, Chin-Shun Hs and Po-Chuang Huang, "Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets," IEEE 2017.
2. Md.Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty, "Agricultural Production Output Prediction Using Supervised Machine Learning Techniques," IEEE 2015.
3. Vinciya, P., and Dr A. Valarmathi., "Agriculture Analysis for Next Generation High Tech Farming in Data Mining" (IJARCSSE) (2016).
4. S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh. "Machine learning approach for forecasting crop yield based on climatic parameters," International Conference on Computer Communication and Informatics (ICCCI -2014), Jan. 03 -05, 2014, Coimbatore, India.
5. M.C.S.Geetha , "A Survey on Data Mining Techniques in Agriculture," Journal of Innovative Research in Computer and Communication Engineering, Feb 2015