CROP PRICE PREDICTION

N. ROSHINI*1, P. GANESH SAI KUMAR*2, P. VENKATESH*3 , G.DHANABALAN *4

*1Student, Computer Science, Vel Tech Rangarajan **Dr. Sagunthala** R&D Institute of Science and Technology, Chennai, Tamil Nadu, India.

*2Student, Computer Science, Vel Tech Rangarajan **Dr. Sagunthala** R&D Institute of Science and Technology, Chennai, Tamil Nadu, India.

*3Student, Computer Science, Vel Tech Rangarajan **Dr. Sagunthala** R&D Institute of Science and Technology, Chennai, Tamil Nadu, India.

*4Associate Professor, Computer Science, Vel Tech Rangarajan **Dr. Sagunthala** R&D Institute of Science and Technology, Chennai, Tamil Nadu, India.

ABSTRACT

This project target is largely concentrated on farming. In farming, farmers play the most important part. When the price falls after the crop, farmers face immense losses. A country's GDP(Gross Domestic Product) is affected by the price oscillations of agrarian products. Crop price estimation and evaluation are done to take an intelligent decision before cultivating a specific type of crop. Predicting the price of a crop will help in taking better opinions which results in minimizing the loss and managing the threat of price oscillations. Price prediction, currently, has come to a veritably important agrarian problem that's to be answered only grounded on the available data. The end of this project is to forecast the crop price for the coming gyration. This work is grounded on chancing suitable data models that help in achieving high delicacy and generality for price prediction. For working on this problem, different Data Mining ways were estimated on different data sets. This work presents a system that uses data analytics ways in order to prognosticate the price of the crop. The proposed system will apply machine literacy algorithms and prognosticate the price of the crop grounded on the Yield and price of the former time's crops. This provides a planter with a perception of the unborn price of the crop that he is going to gather. In this project, we predicted the price of different crops by assaying the former prices of different crops. We used the Random Forest and Linear Regression(Supervised machine literacy algorithms) to dissect the former data and prognosticate the price for the rearmost data and estimate the price of the crop.

Keywords: Price Prediction, Random Forest, Linear Regression, Crop Price, Regression, Forecasting, Machine Learning.

INTRODUCTION

India is a country where farming is one of the top pillars of the economy. The maturity of families are dependent on farming. The country's GDP is primarily concentrated on farming. further than partial of the land is used for farming to meet the requirements of the population of the region. It's necessary to update agrarian practices to meet the demanding conditions. Our exploration aims to break the problem of crop price prediction more effectively to insure growers' inflows. To come up with better results, it uses Machine Learning styles on different data. Productivity can be bettered by understanding and forecast crop prices through this operation. An effective crop price cropping system can offer growers openings that can benefit people in a larger environment. The fast oscillations in crop costs are common within the request. These oscillations in costs are especially owing to the lack of former design. This leads to oscillations in demand and also in the request value of a crop. Once the value rises and growers suffer from an investment loss after the value decreases, it'll lead the crops to be large-priced, getting a disadvantage for consumers. growers aren't apprehensive of the demand within the arising agrarian frugality that's taking place. Machine literacy can be defined as one of the Artificial Intelligence operations that have proven to produce successful forecast models. Growers are no longer looking to use analytics to get data they need to realize practicable perceptivity and make intelligent opinions. utmost of the growers in other countries has started to resettle for automated husbandry. The Random Forest algorithm and Linear Regression belong to the family of learning algorithms that are supervised. The purpose of using a Random Forest is to make a training model that will be used by training introductory decision rules deduced from former data to prognosticate the order or meaning of the target variable. Price prediction, currently, has come to a veritably important agrarian problem that is to be answered only grounded on the available data. The end of this paper is to prognosticate the crop price for the coming gyration. This work is grounded on chancing suitable data models that help in achieving high delicacy and generality for price prediction. For working on this problem, different Data Mining ways were estimated on different data sets. This work presents a system that uses data analytics
ways in order to forecast the price of the crop.

LITERATURE SURVEY

N. Heemageetha (2018)[1]

Crop price prediction using supervised machine learning algorithms created a framework to predict the preparedness of a country to face climate change using a machine learning approach. The study is done for South East Asia. Calculating the predictive index includes data acquisition, training, testing, predicting, validation, and visualization. The study is a precautionary measure to alert the regions and verify their vulnerable index using deep learning.

D. Jayanarayana Reddy [2]
Crop yield prediction using machine learning techniques authors have compared different machine learning algorithms for calculating, Standardised Precipitation Index (SPI) and SPEI. After data collection, Extreme learning methods, Online sequential extreme learning machines, and Self-adaptive evolutionary extreme learning machines. The authors claimed that all three algorithms could be applied successfully to drought forecasting. However, OS-ELM and SADE-ELM perform better than ELM.

Dr. Y. Jeevan Nagandrakumar [3]
An intelligent Crop Price Prediction using a suitable Machine Learning Algorithm has given an analysis of the techniques employed and parameters achieved with limitations that every technique and experiment faced. This paper helps to have a crisp view of Regression Analysis, Linear regression by Sellam (2016). Limitations say that predicting the optimized number of input parameters is more complex.

Mayank Chamapveri [4]
The crop Yield Prediction System for Agriculture Application presented an ample survey of the crossovers between the two areas. A brief study of applications of networking using deep learning techniques is done. We then discuss several techniques and platforms that facilitate the efficient deployment of deep learning onto mobile systems. The authors focus on how deep learning can be useful for mobile and wireless networking. This is a survey paper surfacing the issues and challenges in deep learning in wireless and mobile networks.

Parul Agrawal [5]
Crop Price Prediction and Forecasting System using Supervised Machine Learning Algorithms presents a different perspective to sub-grid parameterizations to a DDA that influences the benefits of high-resolution modeling. Challenges to overcome, but recent advances in computing capabilities and deep learning present novel opportunities that are just beginning to be investigated. Authors believe that machine-learning approaches have huge potential to be explored in connection with the development of a traditional model.

Pranay Malik [6]
Crop Price prediction using Random Forest and Decision Tree Regression have targeted all officials whose main duties include water resources and agricultural management. The final beneficiaries of the output are residents of the area; water users and farmers for whom decision-making can be helped by drought prediction information with finer spatial resolution The models provide spatially distributed detailed drought prediction data of the 6-month Standardized Precipitation Index for the case study area, Fiji. They used Weather Research Forecasting (WRF) model as reference data for overcoming the limitations of a non-dense monitoring network. Also, they used Performance measures of the mean absolute error as well as classification accuracy. The WRF outputs reflect the topography of the area. Hybrid models showed better performance than simply bias-corrected forecasts in most cases. The model based on Extra-Trees trained using the WRF model outputs performed the best in most cases.

PROPOSED WORK

This system aims to overcome the drawbacks of the existing system, A new system based on machine learning algorithms that take the input consisting of the type of crop and eight years of data set is developed. Eighty percent of the data provided is considered for processing and the remaining twenty percent is considered for testing. Linear regression algorithm and Random forest algorithm are used to process the data and it will return output as the predicted crop price. Thus the proposed system will help the farmers to predict the crop price before they sow the specific crop seeds.

- Models can provide farmers, traders, and other stakeholders with valuable information about expected market conditions, enabling them to make more informed decisions about what crops to grow and when to sell them.
- Accurate crop price predictions can help farmers and other stakeholders optimize their operations, reducing waste and maximizing profits.

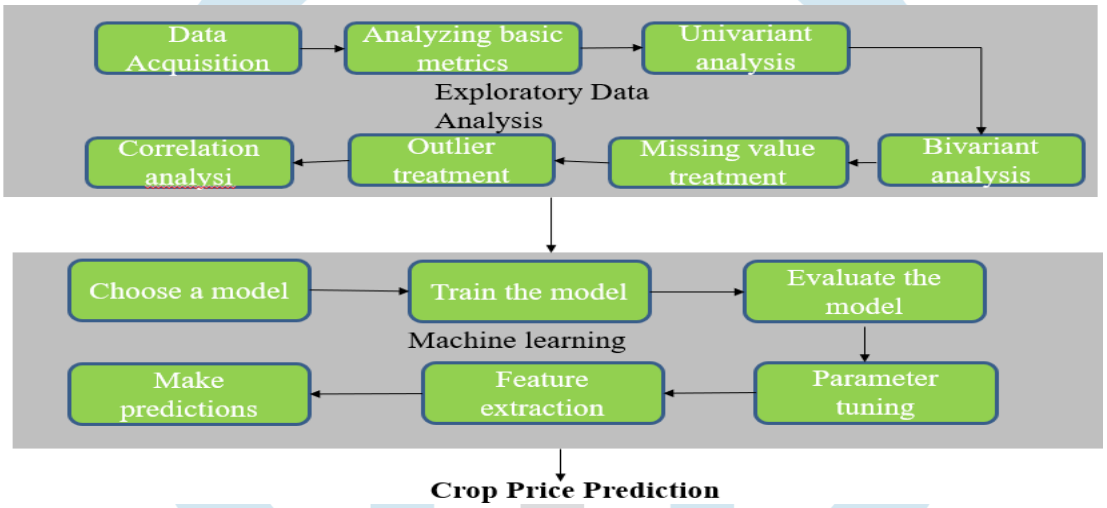
Random Forest Algorithm:
Random Forest is a Supervised Learning set of rules that makes use of ensemble studying strategies for class and regression. Random Forest is a bagging method and now no longer a boosting method. The bushes in random forests run in parallel. There isn't any interplay among those bushes whilst constructing the bushes. It operates with the aid of using building a mess of selection bushes at schooling time and outputting the class this is the mode of the classes (class) or implies prediction (regression) of the person bushes. A random forest is a meta-estimator (i.e. it combines the end result of a couple of predictions) that aggregates many selection bushes, with a few useful modifications:

- The number of features that can be split at each node is limited to some percentage of the total. This ensures that the ensemble model does not rely too heavily on any individual feature, and makes fair use of all potentially predictive features.
- Each tree draws a random sample from the original data set when generating its splits, adding a further element of randomness that prevents over-fitting.

Linear Regression Algorithm:

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between the dependent and independent variables they are considering and the number of independent variables being used. Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hypothesis function for Linear Regression: Y= 1+ 2.x While training the model we are given: x: input training data (uni-variate – one input variable(parameter)) y: labels to data (supervised learning) When training the model – it fits the best line to predict the value of y for a given value of x. The model gets the best regression fit line by finding the best 1 and 2 values. 1: intercept 2: coefficient of x Once we find the best 1 and 2 values, we get the best-fit line. So when we are finally using our model for prediction, it will predict the value of y for the input value of x. Cost function: As mentioned above, we need to identify the best possible values of the parameters based on the given data. To achieve it, first of all, we need to define what best means here. It is the criteria that decide which combination of values of the parameters is better than another combination. It needs to be quantified as well, i.e. how much better a given combination is than another one. This criterion is defined as a cost function. You can consider it as the penalty you pay for a miss prediction or the mistake committed by the model. Once the cost function is arrived at, then the values of parameters that minimize the cost function need to be computed.

Architecture of Crop Price Prediction



The above diagram shows the design methodology of crop price prediction. First, the dataset will be collected and applied for data pre-processing and data cleaning. After the dataset will be applied for exploration and visualization. Then it is split into training and testing data and then applied with Random Forest and linear regression models to predict the crop price.

SYSTEM RESULTS

Out[7]:

	crop	yd1	cost-2016	cost-2017	cost-2018	cost-2019	cost-2020	cost-2021	cost-2022	cost-2023
0	0	4.83	30.36282	68.6016	101.2800	220.77504	134.4600	218.107372	224.491000	160.2860
1	0	7.47	24.30000	50.8700	91.5700	75.95000	95.1300	41.950000	90.850000	31.2000
2	0	9.59	205.05340	68.1100	97.7059	56.80000	76.3276	84.150000	83.630000	12.0100
3	0	6.42	13.13170	201.9000	43.3553	71.69000	28.1800	59.090000	75.970000	35.3000
4	0	8.72	152.23000	211.2719	30.1708	82.65940	84.9082	55.916510	80.651241	40.7807
...	...	...	...	...	...	...	...	...	...	...
425	85	147.00	778.00000	669.0000	807.0000	1015.00000	1031.0000	541.000000	1051.000000	1080.0000
426	85	854.00	1054.00000	985.0000	1951.0000	1943.00000	2424.0000	2516.000000	2476.000000	3625.0000
427	85	856.00	1577.00000	1383.0000	1404.0000	1454.00000	1550.0000	1541.000000	1563.000000	1759.0000
428	85	932.00	1222.00000	1188.0000	1318.0000	1319.00000	1280.0000	1160.000000	1328.000000	1320.0000
429	85	899.00	2443.00000	2132.0000	2589.0000	2729.00000	3319.0000	3531.000000	3700.000000	3455.0000

430 rows × 10 columns

Applied Linear regression with an accuracy of 98%.

```
In [34]: from sklearn.metrics import r2_score
         r2_score(y_test,y_predict)

Out[34]: 0.9820072226671479

In [35]: a=lr.predict([[0.00000000e+00, 8.72000000e+00, 7.22300000e+01,7.82719000e+01, 8.01708000e+01, 8.26594000e+01, 8.49082000e+01, 7.59165100e+01,
                        8.06512411e+01]])

In [36]: a

Out[36]: array([[87.91208897]])

In [37]: b=lr.predict( [[1.00000000e+00, 2.43900000e+01, 1.16300000e+01, 3.23545000e+01, 8.580670000+81, 7.85702000e+01, 6.66794000e+01, 6.00030000e+01,
         8.68739530e+01]])

In [38]: b

Out[38]: array([[112.3953489]])
```

CONCLUSION

We implemented a linear regression model in order to check the accuracy. By applying this Linear Regression model to past data, farmers can predict the price of the crop. The result can be improved by using a large data set. The larger the datasets the better the results are. During the calculation of prices, the role of farmers could be extended and a system can be built which is more transparent and accessible to farmers. Linear Regression and Random Forest models, Linear Regression has the highest accuracy of 98 percent compared to Random Forest with an accuracy of 59 percent Displaying the inserted data set.

REFERENCES

[1] N Heemageetha, ”A survey on Application of Data Mining Techniques to Analyze soil for agricultural purpose”, vol 30, pp.17-18, 2016.

[2] D. Jayanarayana Reddy and Dr. M. Rudrakumar “Crop Yield Prediction Using Machine Learning Algorithm” Supervised learning” vol 12, pp.5, 2021.

[3] Dr. Y. Jeevan Nagandrakumar et.al “Supervised Machine Learning Approach for Crop Yield Prediction in Agriculture Sector” in 2020.

[4] Mayank Chamapveri et.al “Crop Yield Prediction Using Machine Learning” vol 9, pp5-7, 2018.

[5] Parul Agrawal et.al “Crop Yield Prediction Using Machine Learning Algorithm” vol 12, pp.5-6, 2019.

[6] Pranay Malik et.al “Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield using Machine Learning Algorithm” in 2021.

[7] Ramesh Medar et.al “Crop Yield Prediction Using Machine Learning Techniques” vol 11, pp.7-11, 2019.

[8] Shreya V. Bhosale et.al “Crop Yield Prediction Using Data Analytics and Hybrid Approach” vol 3, pp.3, 2018.

[9] Shruthi Mishra, Priyanka Paygude et.al “Use of Data Mining in Crop Yield Prediction” vol 13, pp.7-11, 2018.

[10] Sivanandhini P and Prakash J “Crop Yield Prediction Using Feed Forward and Recurrent NeuralNetwork ” in 2020.

[11] Dr.N.Suma, Sandra Rhea Samson, S.Saranya, G.Shanmugapriya, R. Subhashri, ”1OT Based Smart Agriculture Monitoring System”, Feb 2017.