ESTIMATION OF PRICES IN AGRICULTURAL COMMODITY USING MACHINE LEARNING

CH.Bhanu Kumar[1], Sri.M.Anil[2]

[1]*Student-CSE Department, Sanketika Vidya Parishad Engineering College, Visakhapatnam, Andhra Pradesh*

[2]*Assistant Professor-CSE Department, Sanketika Vidya Parishad Engineering College, Visakhapatnam, AP*

ABSTRACT

India is a country where agriculture and agricultural industries provide the majority of the country's income and economy. Farmers have traditionally had a difficult time predicting prices for agricultural crops. Farmers are currently losing a lot of money owing to price fluctuations caused by climatic change and other price influencing factors. Farmers are unable to obtain the price they desire for their produce. The goal of this project is to develop a decision-making assistance model for agricultural product price prediction. This technique can be used as a guide when deciding what a farmer should plant, taking into account factors such as annual rainfall, WPI, and so forth. The system provides a 12-month forecast in detail. Decision tree regression, a machine learning regression technique, is the methodology we employ in the system.

KEY WORDS: *Price Prediction, Machine Learning, WPI, Decision Tree Regression.*

INTRODUCTION

In our country, agriculture is the most important economic pillar. Agriculture is the main source of income for the majority of the families. To meet the needs of the country's population, 60 percent of the land is used for agriculture. Our study attempts to solve the problem of crop value prediction in an effective manner so that farmers receive guaranteed profits. The market offers a wide range of agricultural items. Agricultural product prices are influenced by a variety of factors, and even the same product can have different pricing in different markets. Agricultural commodity prices are inherently volatile, and they can rise or decrease at any time, wreaking havoc on the economy. This system use the Decision Tree Regression technique to forecast crop value based on data from a verified dataset.

An excellent crop price forecasting system can provide clients with options that will satisfy them to a greater extent. Finally, the results are presented as a web application for farmers to easily access. Farmers will benefit from the work done here to anticipate the costs of horticultural goods since they will be able to sow crops based on their future costs. Regular rates apply to farming things, and these rates are spaced out over the entire year. In the event that these rates are disclosed to farmers ahead of time, it will be guarantying on Rate on Investments (ROI). Horticultural specialists can track these charts and estimate advertising rates which might be told to farmers.

LITERATURE REVIEW
Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector.

The annual agricultural yield is predicted using the Random Forest Algorithm and Decision Tree. It takes data from Kaggle for training purposes. Because of the algorithm's relative simplicity, the system uses fewer computer resources and can provide results quickly. It could only be used with structured data. As a result, it is not uniformly applicable to various types of data.

EPRA International Journal of Research and Development (IJRD)
Volume: 6 | Issue: 9 | September 2021   **- Peer Reviewed Journal**

Crop yield prediction through proximal sensing and machine learning algorithms.

It collects data from New Brunswick and Prince Island using IOT-based devices. Following the collection of data, various machine learning techniques such as Linear Regression, KNN, Elastic- Net, Support Vector Regression, and Linear Regression are used. In comparison to the other methods, the SVR method has the highest accuracy. It was able to explain how internal and external elements such as climate, environment, and other factors influence agricultural output. A tiny dataset was used to test it. It might be more reliable if it's tested on a much larger dataset.

DEVELOPING CROP PRICE FORECASTING SERVICE USING OPEN DATA FROM TAIWAN MARKETS

The performance of the autoregressive integrated moving average (ARIMA), partial least square (PLS), artificial neural network (ANN), and PLS combined with response surface methodology (RSM) termed RSMPLS was implemented and compared. The non-linear relationship between past prices can be induced using RSMPLS. The developed service is also linked into the smart agro management platform, offering an interface for retrieving historical prices and forecasting future prices. Only one parameter is used in the model (prices). More variables like as climate, market location, and planting space, will yield more accurate findings.

Demand based crop recommender system for farmers.

The Decision Tree algorithm is used to classify and recommend crops for farmers to cultivate. To make the results more relevant, the dataset only included data from the previous year. They also created a user-friendly GUI with nlp algorithms for text-to-speech conversion and vice versa. Farmers will benefit greatly from the GUI that contains nlp algorithms. The results will be more relevant because the dataset will continually be updated to the most recent year's data. There are insufficient factors to consider when determining the crop. More characteristics are required to obtain more accurate findings.

Decision Making Support System for Prediction of Prices in Agricultural Commodity

A decision-making assistance model has been created to assist farmers in forecasting prices. This concept includes a portal where farmers must check in with their credentials to access their account. Farmers must input the commodity name as well as the crop's prior selling price. This programme will be able to provide typical pricing for a certain crop based on prior prices, which would help farmers make better decisions and forecast prices. They are using inefficient data to anticipate crop prices because they are using previously logged in prices by consumers and essentially taking an average. Because no additional parameter is taken into account, this model will not always produce accurate and exact results.

An efficient analysis of crop yield prediction using Hadoop framework based on random forest approach.

On the Hadoop system, the Random Forest Algorithm is combined with MapReduce programming. Will improve the scale of the process by providing efficiency in managing massive amounts of data. For crop yield prediction challenges, there may not be a modular and platform- independent solution.

METHODOLOGY

Tools used: Python and flask are the two most important needs. Python is a high-level programming language in which the system models are defined. As a result, you'll need a fully functional Python 3.5+ environment including the sklearn, numpy, and pandas libraries. Another requirement is the flask html-python framework. Flask is a web framework that was created in Python. It doesn't have its own set of tools or library dependencies.

The website was created to allow users to access and view the algorithm's results. The following techniques are being utilized to create the most interactive website possible, where users may learn about the quantity of land accessible, the amount of land used for cultivation, the number of farmers who use the land, and the many sorts of crops that have been farmed. Farmers can also communicate with one another by using the social media buttons at the bottom of the website. They can communicate with agricultural experts by calling the helpline number or visiting the website; there is also a mobile application that they can use to get their questions answered. This is a system that assists a farmer in knowing what is going on in the present and predicting what will happen in the future, making it easier to make decisions.

Process: The current study focuses on supervised learning models, which are part of machine learning. The conceptual model is defined in distinct structures and multiple viewpoints of the system by the system architecture. The crop price can be determined using the proposed framework. This model depicts a large

---

volume of data being captured and preprocessed to remove undesired datasuch as NULL and outlandish values. We divided the dataset into training and testing datasets during the preprocessing procedure. Using appropriate supervised learning methods, train the dataset to detect the crop price existing in the dataset. Apply machine learning algorithms to find crop prices for each new data that has appeared in the dataset. Following data collecting, a suitable machine learning algorithm must be used to compute the model's efficiency and capabilities; we used a variety of machine learning algorithms, including random forest, polynomial regression, and decision trees. For the suggested model, accuracy and precision will be calculated. The flow data, machine learning approaches, and modules for detecting crop pricing and feature selection modules are all part of this system architecture.

Fig 1. System Architecture
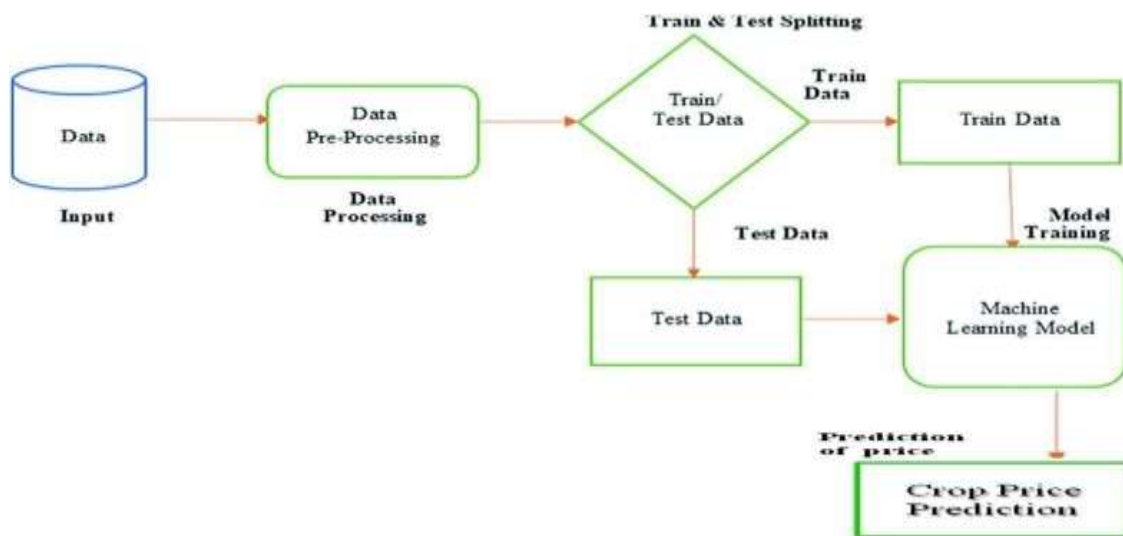
The following implementation modules are a big part of the research work.
1. Data Gathering
2. Exploration of Data
3. Prediction using machine learning
4. Web-based software

Data Gathering
The crop data was gathered from a public repository and used to create the dataset. There are a few datasets that provide information. We were able to get data containing the specifics of the individual crops rainfall.

Exploration of Data
Data Exploration is a type of exploratory data analysis. Data analysis (EDA) is a critical step that occurs when a component is gathering and analysing data, and it must be completed before any demonstrating. This is due to the fact that it is critical for an associate in nursing data research worker to virtually comprehend the data's concept while avoiding producing suspicions. The influence of knowledge inquiry is frequently extremely beneficial in gaining a handle on the structure of data, the appropriation of qualities, and thus the distance of exceptional characteristics and interrelationships inside the informative index.

Prediction using machine learning
One of the most widely used and useful models for supervised learning is the Decision Tree. It can be used to tackle both regression and classification problems, albeit the latter is more widely utilized. There are three sorts of nodes in this tree-structured classifier. The Root Node is the first node in the graph, and it represents the complete sample. It can be further divided into nodes. The features of a data collection are represented by the interior nodes, while the decision rules are

EPRA International Journal of Research and Development (IJRD)

Volume: 6 | Issue: 9 | September 2021   - Peer Reviewed Journal

---

represented by the branches. Finally, the outcome is represented by the Leaf Nodes. This algorithm is quite useful for resolving decision-making issues.
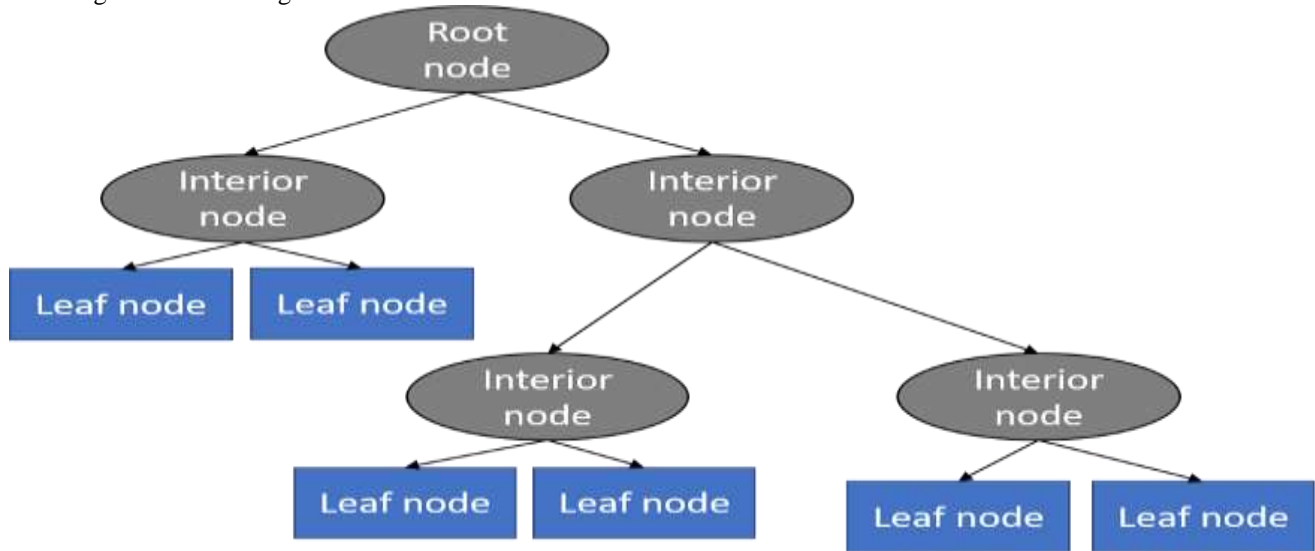

Fig 2. Decision tree model

A specific data point is traversed through the entire tree by answering True/False questions until it reaches the leaf node. The average of the dependent variable's value in that particular leaf node is the final forecast. The Tree is able to estimate an appropriate value for the data point after numerous iterations.

STEPS TO IMPLEMENT THE ALGORITHM

Step 1:- Produce a dataset with training data for rainfall and the wholesale pricing index. **Step 2:-**Select all the rows and column 1from dataset to "x" Which is independent variable **Step 3:-**Select all of the rows and column 2 from dataset to "y" Which is dependent variable **Step 4:-** Fit decision tree regress or to the dataset
Step 5:-Predict the new value
Step 6:-Visualize the result and check the accuracy

RESULT AND DISCUSSION

The decision tree is one of the most powerful complete regression algorithms available, capable of working successfully and efficiently on tiny datasets. The tree is built using the criteria function. The entropy, often known as the impurity measure, is one of the criteria employed. Entropy is the most popular impurity measure or splitting condition in classification. It is defined by:

Here, p(i|t) is the proportion of the samples that belong to class c for a particular node t. The entropy is therefore 0 if all samples at a node belong to the same class, and the entropy is maximal if

$$E(t) = -\sum_{i=1}^{c} p(i|t) \log_2 p(i|t)$$

we have a uniform class distribution.
Three equations are used to evaluate the performance of decision tree regression:
1. Mean absolute error   2. The squared error of the error3. R2 score
The above three equations are used to calculate the accuracy of regression methods. Different algorithms were created and tested, and the results were compared to determine which one was the best. Although both the choice tree regression and the random forest algorithms performed well, the decision forest approach produced the best results for the majority of crops. The random forest algorithm, on the other hand, only worked for a few algorithms. As a result, decision tree regression is utilised to anticipate prices, which aids farmers in making decisions with a 95.4 percent accuracy.
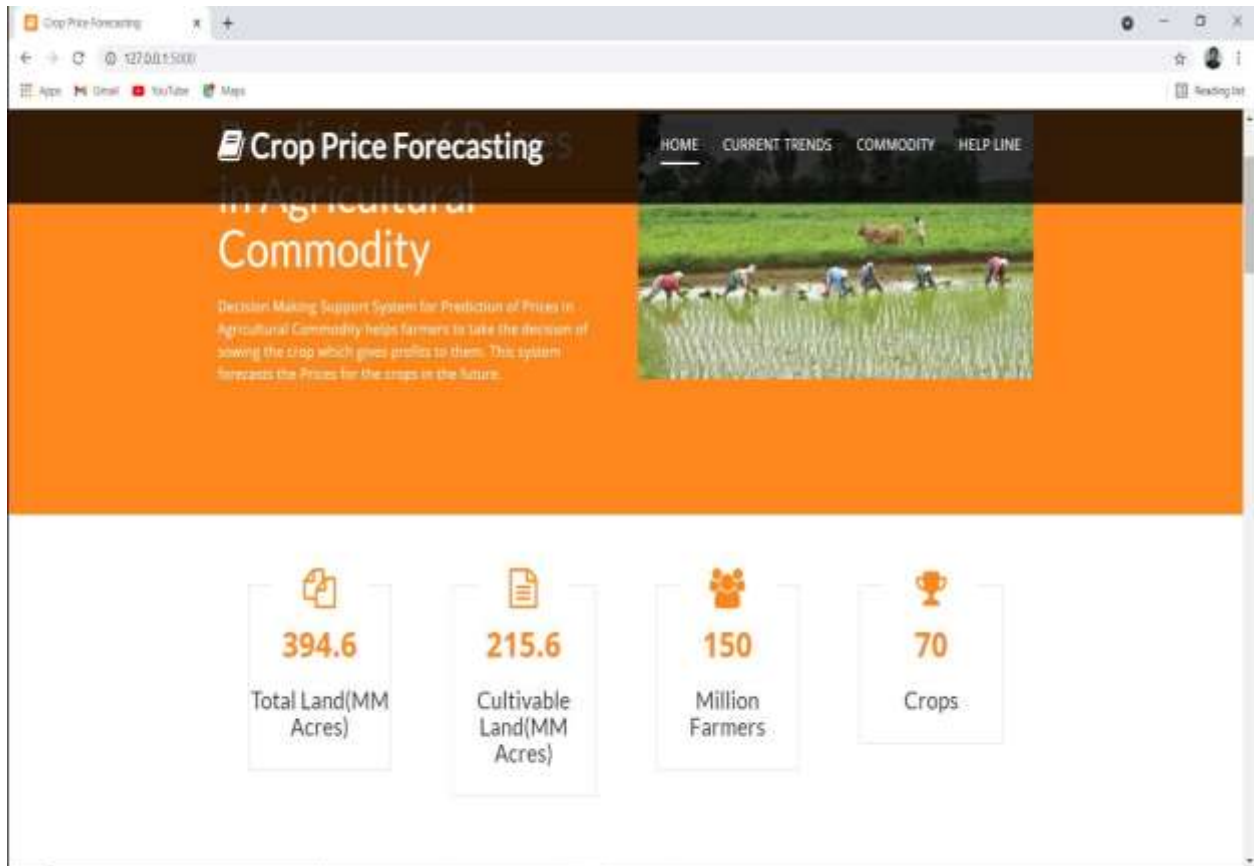
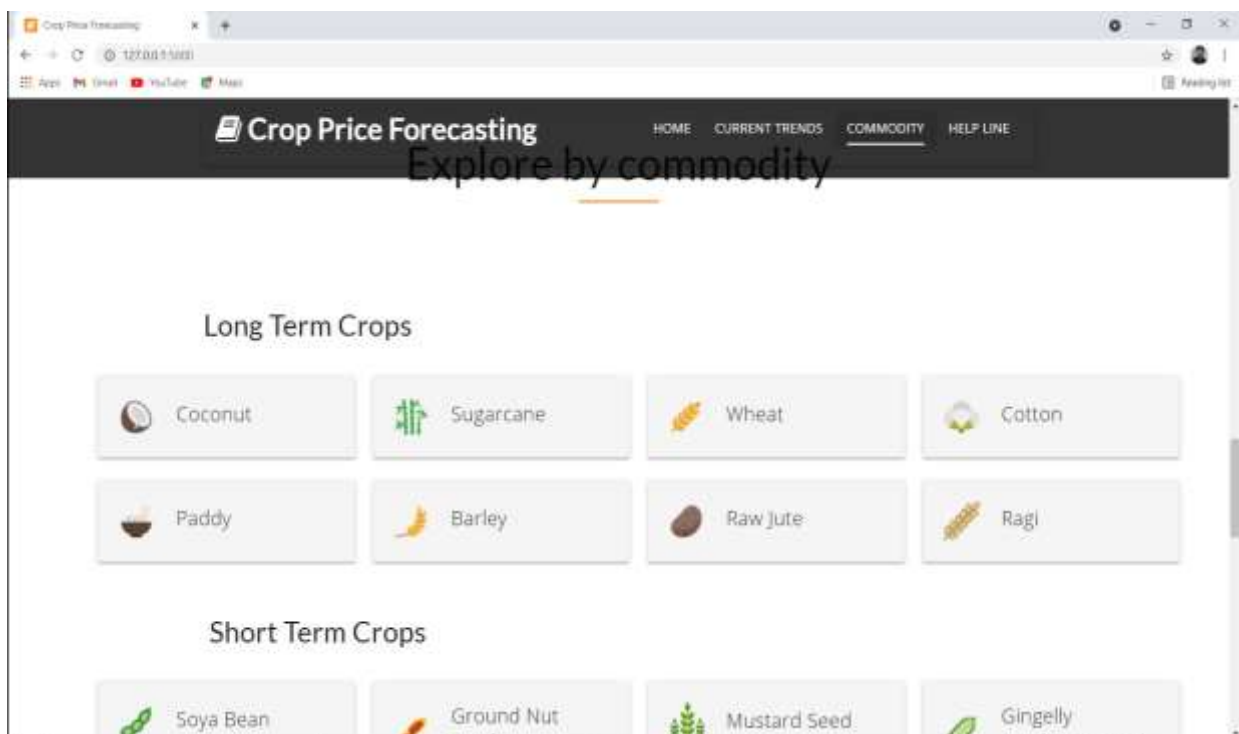**Fig 3. Home page of the web site with the amount land and crops information.**

Fig 4. Long Term and Shot term crops list are displayed from which users can select a crop.
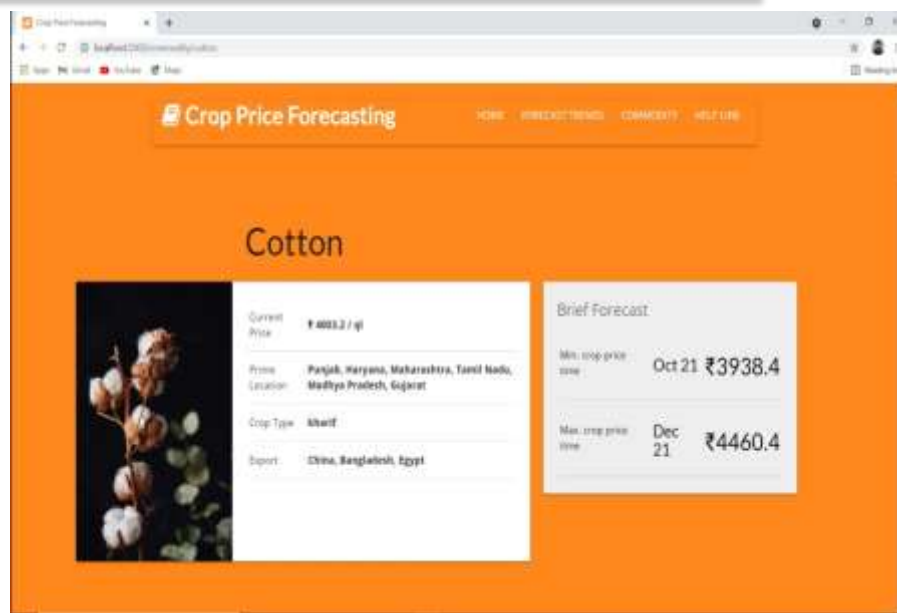
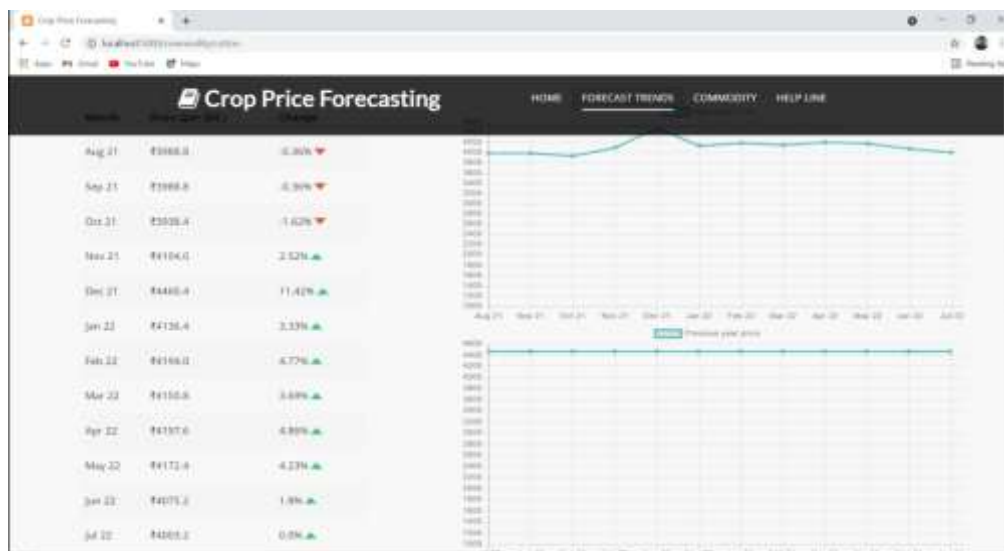**Fig 5. The crops details like minimum and maximum price and location were displayed.**

Fig 6. The forecasting trends of the crop like its next 12 months price and the graphs that shows the statstics of the price for present and next 12 months.
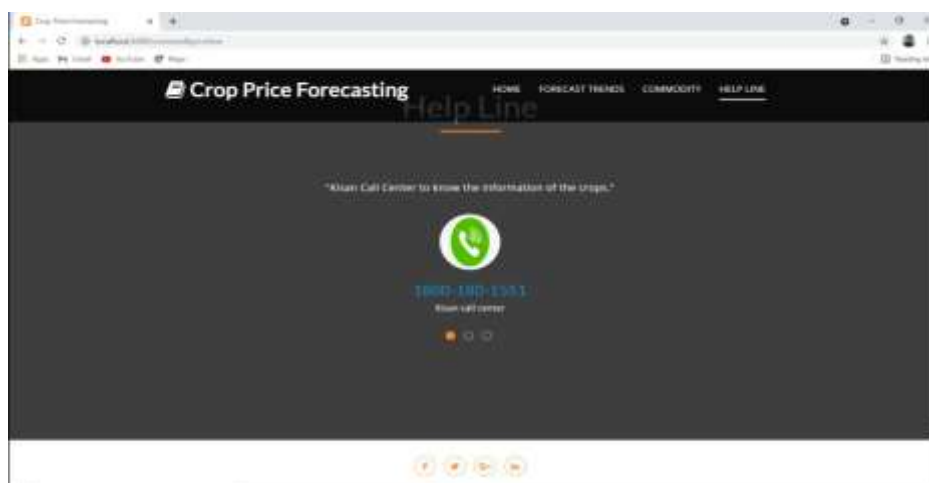

Fig 7. The help line details and the social media links were provided here.

EPRA International Journal of Research and Development (IJRD)

The user does not provide any input to the model. It simply displays information such as current trends, with the top lucrative crops and the top non-profitable crops listed. It also contains information on the most important commodity crops. The crops can be explored separately, with a list of long-term and short-term crops provided. Farmers can select the crop then they can navigate to the comodity and can view the details like maximum and minimum price per quinta and the area of growth and the locations to which the products are been exported. They can observe the 12 months forecast beside the graphs which shows the statistics for present year and the next year.

CONCLUSION

The proposed model was created to assist farmers in making better decisions about which crop to plant and when to plant it, as well as where to plant it. Our technology forecasts the price of the crop of choice, providing vital information to the farmer even before the growing process begins. For price prediction, a variety of algorithms such as decision trees, neural networks, and SVM can be utilised. A decision tree is used in our model. It has been trained on a variety of kharif and ragi crops (such as paddy, arhar, bhajra, barley, and others) and has a high level of accuracy.

The project uses a web application to estimate crop prices and forecasts, and it is based on efficient machine learning techniques and technologies with a user-friendly interface. The training datasets gathered thus far give sufficient information for predicting market price and demand. For each technique, the root mean square error is determined to properly measure the accuracy of each system used, and the most accurate system is then chosen. This approach will enable farmers to make better decisions when it comes to bidding for better market prices for their crops. As a result, the approach assists farmers in lowering their hardships and preventing suicide attempts.

REFERENCES

1. https://www.educba.com/machine-learning-algorithms/
2. Kaur, Manpreet, Heena Gulati, and Harish Kundra. "Data mining in Agriculture on crop price prediction: Techniques and Applications." International Journal of Computer Applications 99.12 (2014): 1-3.
3. Ramesh, D., & Vardhan, B. V. (2013). Data mining techniques and applications to agricultural yield data. International journal of advanced research in computer and communication engineering, 2(9), 3477-3480.
4. 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.
5. Sadiq A Mulla, S. A. Quadri. "Crop-yield and Price Forecasting using Machine Learning." International journal of analytical and experimental modal analysis. August 2020.
6. 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.
7. 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.
8. T. Siddique, D. Barua, Z. Ferdous and A. Chakrabarty, "Automated farming prediction," 2017 Intelligent Systems Conference (IntelliSys), London, 2017, pp. 757-763.
9. 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.
10. 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