IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

FARM-PRODUCTS PRICE PREDICTION USING MACHINE LEARNING

K. Chanikya, D.Anusha, A.Rakesh

Department of Electronics and Communication Engineering, Sreenidhi Institute of Science and Technology, Hyderabad

ABSTRACT

India is an agriculturally based nation, and farmers give their entire lives to the country's economy. They encounter serious issues when the crop is not worth the price and when they are ignorant about the marketing price for the crop or commodity in question in a given cluster. Farmers currently suffer significant losses as a result of price volatility brought on by climatic change and other price-influencing factors. From the standpoint of the agricultural industry, the current stage of demand for a crop or Agri-product is reflected in its market price. In order to adapt the production schedule and maximize profit, it is crucial for Agri management to simultaneously monitor and estimate market prices. Accurate crop price forecasting is crucial for crop production control.

Such a forecast will assist related industries in planning their supply chain strategies. Generally speaking, farmers can boost their profit and avoid severe loss by using this application, which provides them with a forecast in advance. which ultimately boosts the economy of the nation. Therefore, the goal of this project is to create an application that can calculate the price of agricultural products and assist farmers in anticipating the price of a given commodity. This is possible with the aid of machine learning algorithms, and it primarily relies on the discovery of pertinent data models that aid in reaching high accuracy and generality for price prediction.

INTRODUCTION

India is an agriculturally-based nation, and the farmers' community is the foundation of the industry. Farmers have a difficult time deciding which crops to cultivate in the modern world with all the changes that are taking on. The agricultural market environment is changing in an unprecedented number of ways both locally and globally. These elements have an effect on agriculture earnings as well as agricultural prices. The majority of farmers in rural areas are unable to understand and interpret price and market behaviour in their benefit. Therefore, market knowledge and intelligence are essential to help farmers and traders decide what to raise, when to sell it, and where to sell it. The price knowledge that a farmer needs in the current environment is the most crucial marketing information input. Most farmers still don't understand these problems very well. Forecasting crop prices helps farmers plan their subsequent crop while preventing hyperinflation. The effectiveness of various models has been investigated and contrasted. Agricultural goods are subject to seasonal rates; these rates are yearly averages. The goal of Crop Value Forecasting using Decision Tree Regressor is to effectively address the crop value forecast issue so that poor farmers can benefit from certain advantages. India is a country where 52% of the population is engaged in farming out of which 82% farmers are small and marginal, It is only our obligation to offer a centralised method to address the issue of erratic crop value predictions and to ensure that our farmers have the same opportunities.. Our model will contain data of around 22 commodities (including all kind of crops). We prefer to develop a more efficient approach that will more accurately predict crop value than present methods. Also, by providing this application, the farmer can become more technologically savvy, allowing him to be a more active participant in this smart age. This will aid in the expansion of any country's economy, particularly India's, as agriculture is the country's primary source of income. Any country's economy will benefit from this, but India's would benefit the most as agriculture is the main source of income here.

PROBLEM STATEMENT

One of the biggest hazards that a farmer now faces is price volatility. Nevertheless, the farmer has few options for reducing this danger. In the modern world, price volatility is a product of both global (trade) dynamics as well as domestic supply and demand. Accurate pricing estimates can help farmers in this situation anticipate price changes and arrange their marketing appropriately. Therefore, using various input data, we will design, develop, and implement the training model in this project. In order for the machine to use machine learning

concepts to extract the crop's predicted price from the data and learn its attributes..

SOLUTION

The proposed system aims to estimate or predict the price of agricultural products using the collected data from past years to train the model for the project. The strategy uses machine learning to build a predictive model that takes into account a number of variables, including the year, month, rainfall, and WPI. In this research, we make use of machine learning techniques, particularly Decision Tree.

Challenges are a significant issue that will negatively affect the current endeavour. Some of the challenges that the crop price forecast system may encounter include the following:

• Choosing the right dataset and then fine-tuning the parameters to make the project more effective in achieving the intended results.

• The error rate increases as the environment changes in real time; the model must be developed with reduced processing efficiency and power in mind.

WORK FLOW

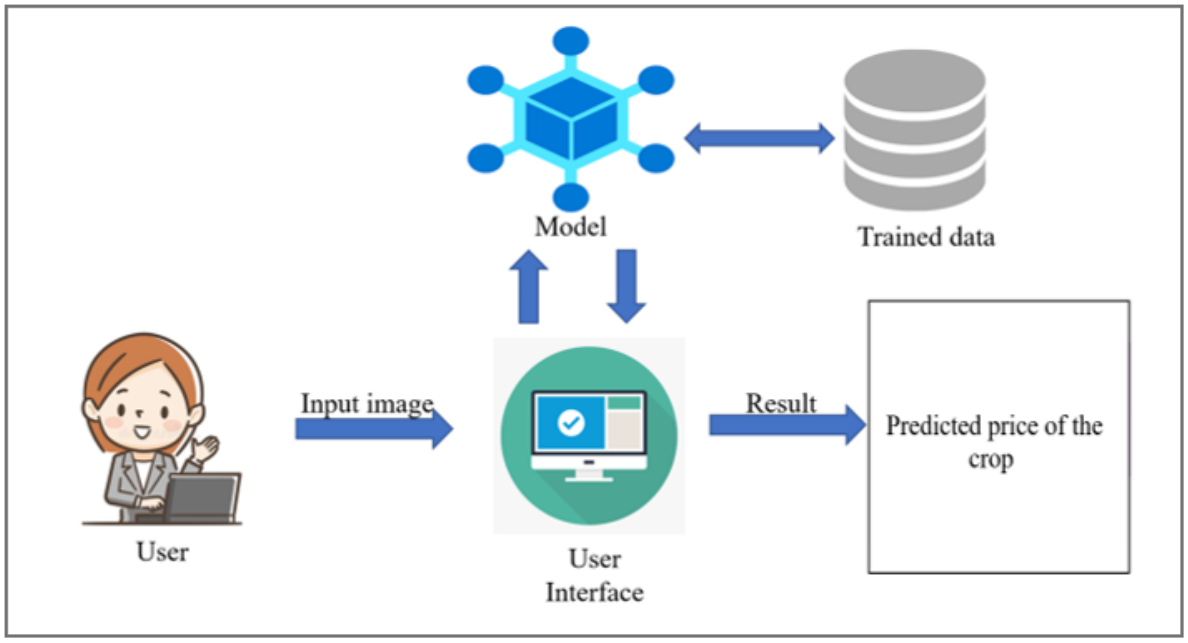

Fig. 1. Project Workflow

EXISTING METHOD

The development of an arecanut price prediction model. The performance of the traditional time-series models Holt-Seasonal Winter's Method, SARIMA model, and Machine learning model LSTM were investigated in order to identify the most frugal model. An exploratory investigation of the price time-series data found that trends and seasonality were present. The purpose of the research (in this paper) is to determine whether a

machine learning model can be used to simulate and predict commodity futures market values, and if so, whether the application is sound and effective. The main application of machine learning technology is to forecast the price of agricultural futures based on an examination of the fundamental factors influencing the market.

This study examined the effectiveness of ANNs for short-term price forecasting of agricultural commodities. They are compared to futures prices and conventional ARIMA models as a benchmark. Additionally, they assessed if utilising a hybrid model is more beneficial than using a single forecast approach, as suggested by the available literature. Their findings are ambiguous.

Prices for crops can vary greatly, especially for annual and tree commodities. Unpredictable calamities like droughts and floods can lower the value of agricultural products, which has an effect on the entire market and results in considerable losses for farmers, suppliers, exporters, and stakeholders. The weighted moving average model is shown to have been utilised to estimate coffee demand and share price in the Indian coffee supply chain stock market, enabling exporters and stakeholders to act quickly to reduce financial loss and maximise profit.

Farmers are severely impacted by bad crop years when a considerable quantity of debt is incurred. Based on a projection, the government and insurance firms can assist farmers financially by offering insurance coverage and low-interest loans. With the help of multiple time-series, machine learning, and deep learning models, the prices of agricultural commodities are predicted.

Different prediction techniques were combined to create a single-layer neural network that was trained to estimate exchange rates more accurately.

The study report suggested a fusion model that included HMM, artificial neural networks, and genetic algorithms to forecast how the financial markets would behave and how much money would change hands the following day.

The study report compared the performance of ANN and ARIMA models in predicting gold futures prices, concluding that ANN outperformed ARIMA by a wide margin. It also examined the factors that contributed to these results.

Applications that use Linear Regression, a supervised Machine Learning technique, may predict the price of a single item in the future. But these models' precision is too low. The accuracy rates of the current system are unsatisfactory.

PROPOSED METHOD

The suggested method focuses on creating an intuitive application that allows us to forecast the future pricing of the crop that the user has chosen. Algorithms for machine learning can be used to do this. To develop the model that aids in forecasting the price for the chosen product, we are utilising the Decision Tree Regression algorithm, one of the supervised machine learning techniques. There are a few phases that must be taken in order to construct the model, and they are as follows:

• Gathering the necessary data is the initial stage. You must choose the crops whose prices you plan to forecast.

• Once the data has been gathered, it must go through pre-processing, one of the most important processes in developing the project. Preprocessing is required to get rid of any unnecessary information in the data that won't affect the outcome.

• The preprocessed or cleaned data is saved after data cleaning.

• Using the regression algorithm, namely the Decision Tree Regression, this data is being used to generate the model.

• The Decision Tree Regression model is then fitted using the train and test data. Once the model has been fitted, we can use it to forecast the price of the chosen crop.

• The Graphical User Interface is then connected with this model to offer the user and system interaction.

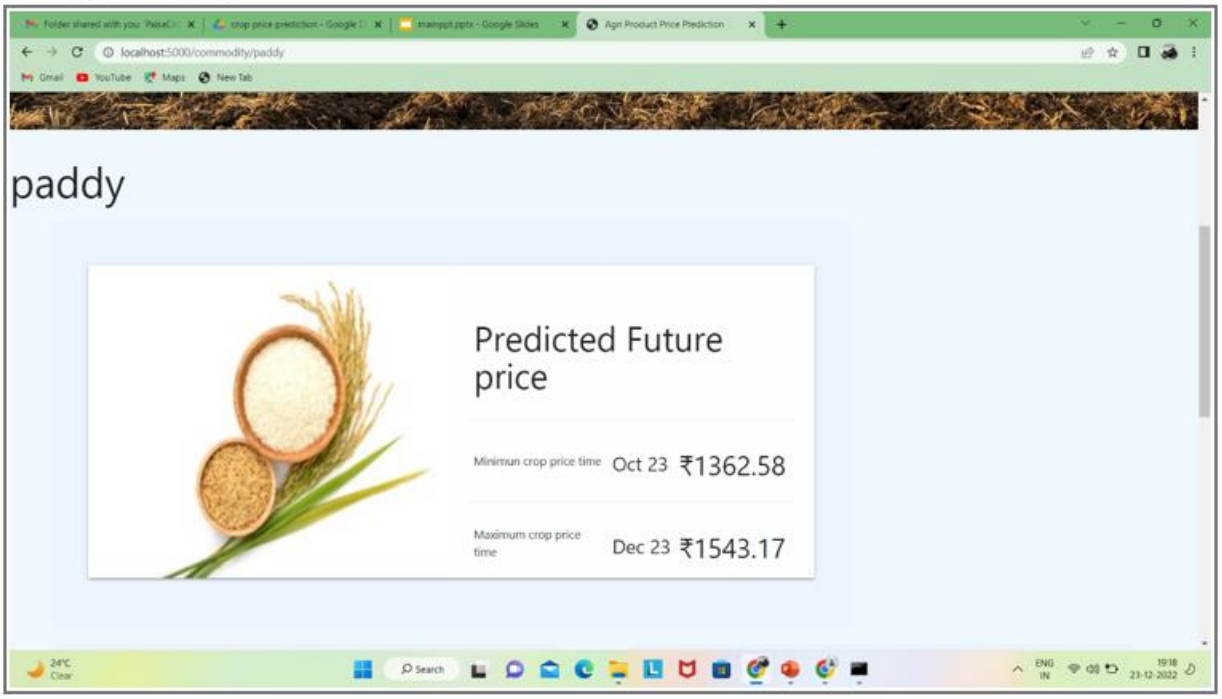

Fig. 2. End output Image

TECHNOLOGIES USED

MACHINE LEARNING

With the use of machine learning (ML), which is a form of artificial intelligence (AI), software programmers can predict outcomes more accurately without having to be explicitly instructed to do so. In order to forecast new output values, machine learning algorithms use historical data as input. The input data, the trained model, and the output are all part of the proposed system's architecture. The trained model processes the feature values that the user supplies as input. Additionally, the user sees the output via the interface. Training after which performance review will follow. As a result, we obtain the trained model's results. We can predict the outcome after providing the input.

- Data pre-processing
- Data standardization
- Data Cleaning
- Handling missing data
- Handling null values
- Handing duplicate values
- Managing the outliers

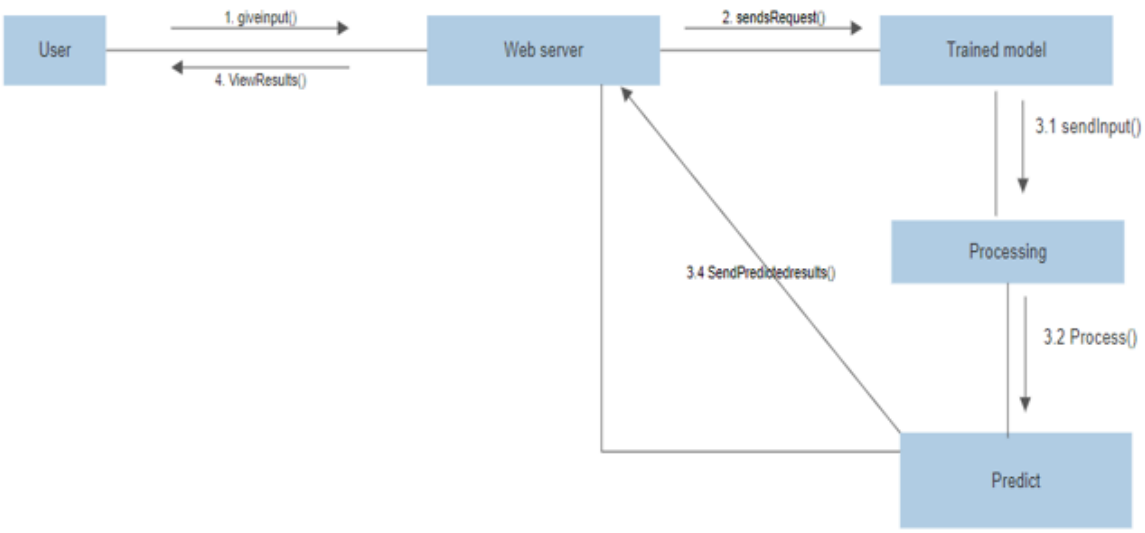

Fig. 3. Collaboration diagram.

SOFTWARE REQUIREMENTS

Python and Flask are the primary needs. The system models are defined in Python, a high level language. Therefore, a fully functional Python 3.5+ environment with the libraries Skearn, Numpy, and Pandas is needed.

Another essential component is the flask html-python framework. A web framework known as Flask was created in Python. It doesn't have any unique requirements for tools or libraries.

Additionally, it is a micro framework with a backend. For the front end of the website to be created, HTML and CSS are needed. Therefore, the prerequisites are Python, HTML, and CSS for the front end.

- Programming Language : Python
- IDE : Visual Studio Code
- Front end :HTML,CSS
- UML Design : Start UML
- Tools : PIP
- Framework : Flask

How Machine Learning works?

Data pre-processing:

Data pre-processing is the most crucial component of every machine learning-based system. The obtained dataset can be in raw format and cannot be put into the machine directly. To make them more accurate and useful, nearly all of the datasets need to be cleaned, processed, and converted. This is so because a raw dataset by definition contains null values, duplicates, outliers, and missing values. Such a dataset produces disastrously inaccurate findings when used without processing. Since every dataset is unique, we may encounter a variety of difficulties. Depending on the technology being used, there can be several types of data that need to be homogenised. In the case of our model, the missing values, duplicate values and the null values are removed in the data pre-processing. Outliers are also eliminated.

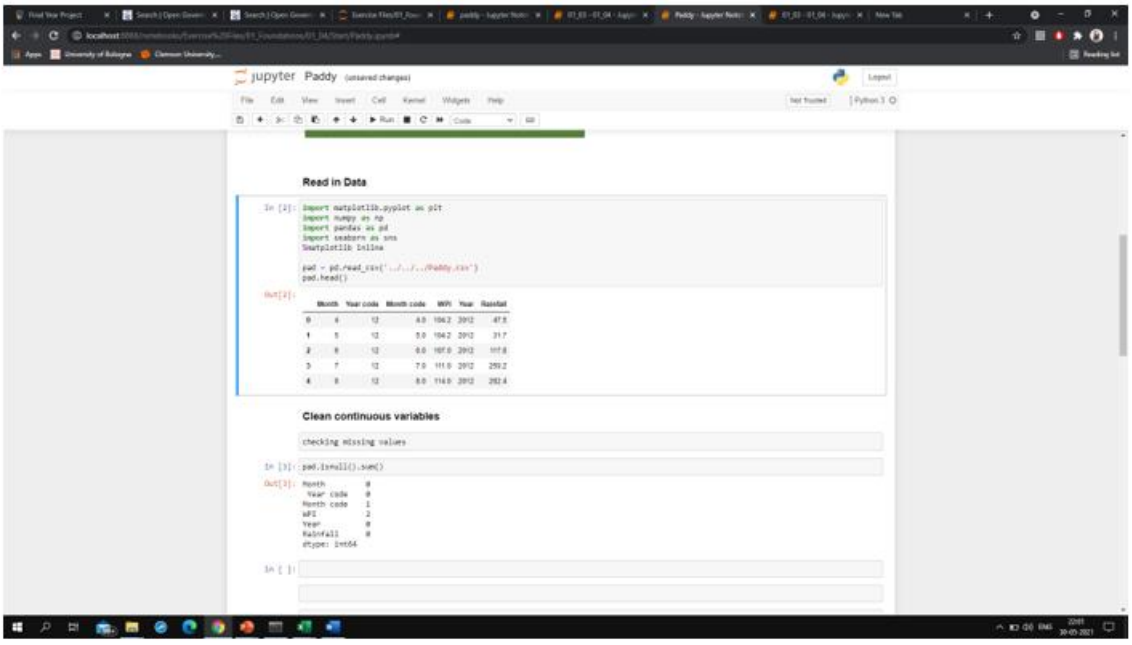

Fig. 4. Reading Data

Data standardization:

Data standardization, which aids in removing inconsistent data, is another essential data pre-processing approach. The internal consistency of the data is improved by the data standardization procedure. It is useful for converting characteristics with a Gaussian distribution into a Gaussian Standard Gaussian Distribution that differ in means standard deviation. Using the scikit-learn library, we standardize the data in the dataset for our project.

Data cleaning:

One of the essential elements to achieving strong machine learning accuracy is data cleaning. Additionally, it is a crucial stage in our project. Despite its critical necessity, data cleaning is typically a process that is not emphasised by most specialists. It aids in locating and eliminating missing values, outliers, and inconsistent data that could obstruct the model's proper execution. In general, even when employing a simple method, we may get better results if we are using better, cleaner data. Our dataset includes various properties of various sorts. As a result, different techniques would be needed to clean the dataset.

The following list of data cleaning methods is an explanation of the ones we utilised in our project:

• Handling missing data: In the realm of machine learning, missing data can be a deceptively difficult problem. These could result in incorrect interpretation and outcomes. We can't just delete the record to solve this problem. In order to ensure that the critical records are not overlooked, we must find a strategy that would assist in filling in these missing values. We approach this problem in two ways. The first method involves eliminating the entries for the missing data, whereas the second method involves filling it in based on prior observations. If the data is of lower quality or is insignificant, the first step might be carried out. This approach might not be advised because crucial information would be lost.

• Handling null values: Since null values have the same problem as missing values, it is possible to avoid them by either deleting the entry or filling it in with the proper statistical procedure. In our project, we dealt with null values by using the mean of the attribute column or the prior entries in or. To carry out the statistical operations, we made use of a suitable technique from the Python library.

• Eliminating duplicates: Our model, or any other machine learning model, would be significantly impacted by the presence of duplicate entries in the dataset. Therefore, the duplication problem needs to be fixed for the model to run successfully. In our project, we eliminated duplicates from our dataset using the drop duplicates approach. The next sections go into detail about this approach.

• Handling the outliers: Handling the outliers presents another difficulty in creating a machine learning model. For instance, if a model contains outliers, linear regression may perform less quickly or accurately than a decision-tree. Unreasonable or improbable data are not appropriate for the dataset. For instance, we are unable to include entries with ages of 300 because doing so would be practically impossible and would harm our model. Therefore, eliminating the outliers and working on the project is a better plan of action. In order to discover and locate the outliers for the qualities in our project, we produced various graphs on the attributes. . Significantly we also transformed and modified the data for improving our dataset. Hence, we applied the methods available in the python library to study the data in the dataset so that we could identify any abnormal entry or anomalies in the data. Python libraries used have provided us with appropriate methods to visualize the data and detect the missing values and outliers and overcome the problems in our dataset.

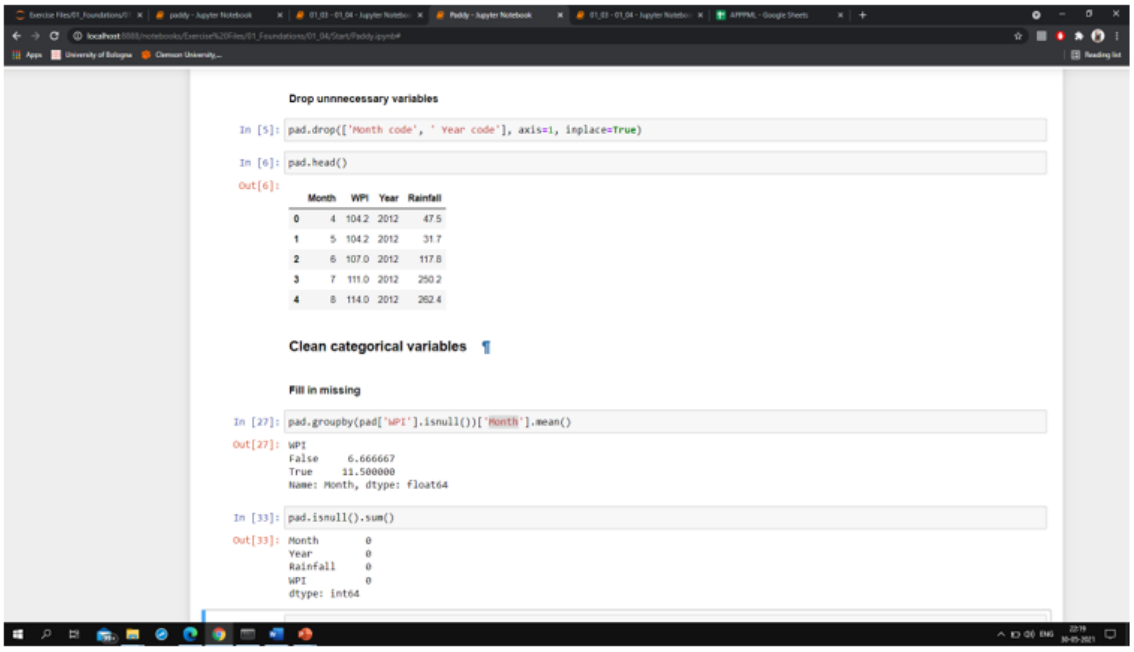

Fig. 5. Cleaning Data.

LANGUAGE USED

We wrote the python code for the system on a jupyter notebook in order to implement it. An open-source web tool called Jupyter notebook makes it easier and faster for programmers to edit and write code. The Jupyter Notebook helps a variety of programmers use Python in large projects more effectively. Numerous libraries, including Sklearn, Numpy, and Pandas, are supported. Having the ability to run code block by block is one of Jupyter Notebook's key benefits. This facilitates the debugging process and improves error handling. HTML and CSS are also used.

Our system makes use of several libraries, including Pandas, Numpy, Sklearn, and Flask. The offer a number of

benefits for the machine learning process. For the Python programming language, the Pandas package, for example, offers simple-to-use data structures with great performance and a tool for data analysis.

HTML

The markup language used to create websites and web applications is called Hypertext Markup Language (HTML). Together with JavaScript and Cascading Style Sheets (CSS), it is the third member of the trinity of foundational technologies for the World Wide Web. HTML files are downloaded from a web server or locally saved files by web browsers, who then transform them into multimedia web pages. Initially, HTML included visual clues for the document's appearance and logically represented the structure of a web page. The elements that make up HTML pages are known as HTML elements. Using HTML techniques, images and other objects, such as interactive forms, can be incorporated in the generated page. By specifying structural semantics for text elements like headers, paragraphs, lists, links, quotations, and other elements, HTML enables you to create well-organized texts.

CSS

Developers choose colours, attractive fonts, and various layouts for HTML pages in order to make them appealing to consumers. The CSS is in charge of all of this effort. To put it simply, CSS is used to style an HTML document. Its architecture allows information and presentation to be separated, making it easy to alter the content without affecting the design. Additionally, it permits sharing a single CSS file for styling across several web pages, saving time and effort.

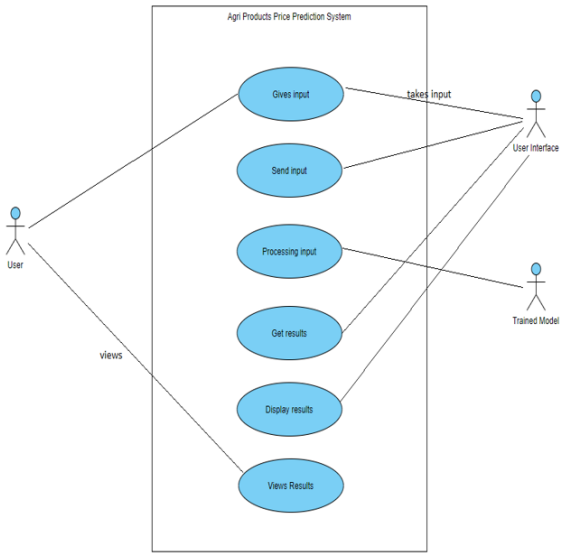

Fig. 6. Use case Diagram

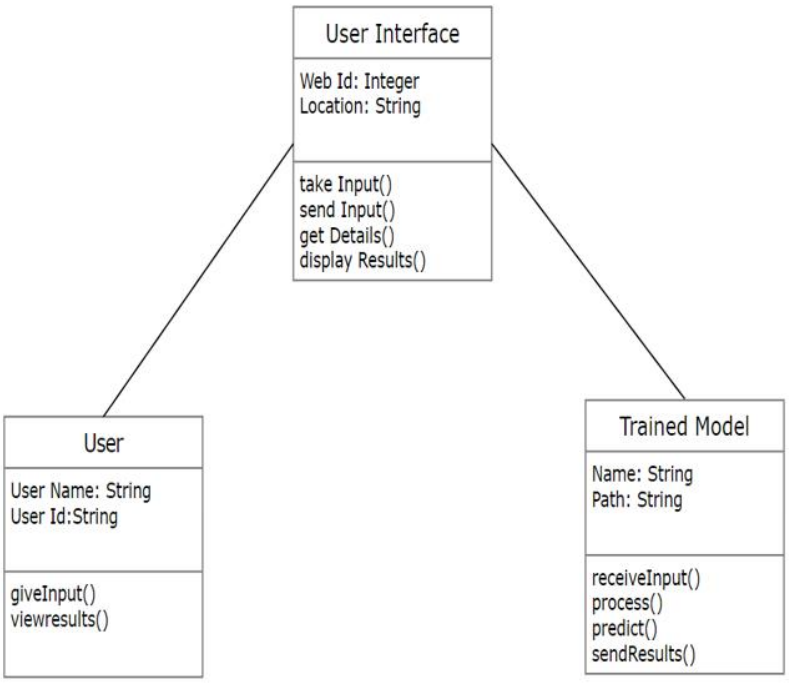

Fig. 7. Class Diagram

IMPLEMENTATION

Exploratory Data Analysis

This is the stage of the procedure that is both more significant and interesting. The step of exploratory data analysis in design thinking is common. In this step, we carefully examine the data to find any hidden insights. It is referred known as the machine learning brainstorming stage for this reason. In order to acquire insights and discover various correlations between the variables, we use a variety of data visualisation techniques in the form of graphs and plots in this step. This could aid in finding any data imbalances and also help to decrease the data to make it more useful. To achieve this, we make use of the Python library functions.

Algorithms Implemented

The project's major goal is to find an algorithm that offers greater accuracy than the current system. We employ various machine learning models in this situation and determine the algorithm with the highest accuracy. The model is subsequently subjected to a performance evaluation, and in response, we create a final trained model. The algorithm is then used to deploy the model and predict prices. The algorithm we used in our system is listed below.

Decision Tree

The classification procedure known as the decision tree is employed in both classification and regression analyses.

A decision tree is a supervised machine learning model that anticipates a goal by learning choice rules from features. They are constructed recursively, with the child nodes coming last after the root node. Both the model hyper parameters and the data properties help with node formation.

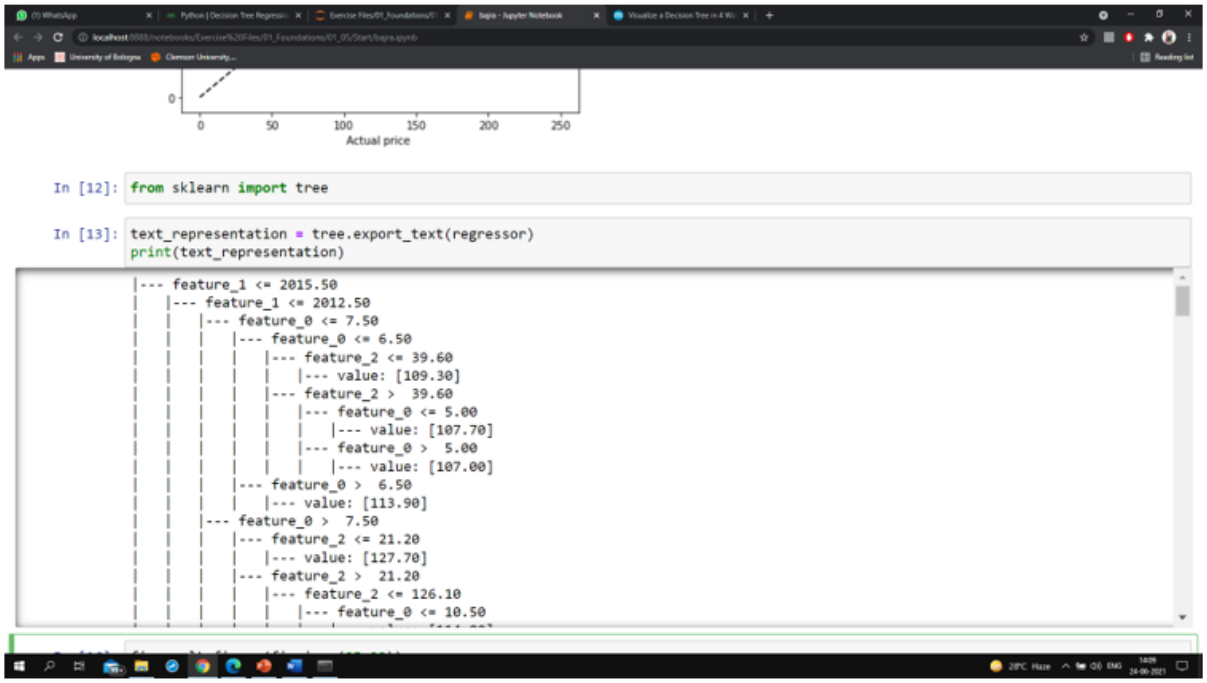

Fig. 8.1. Representation of DT

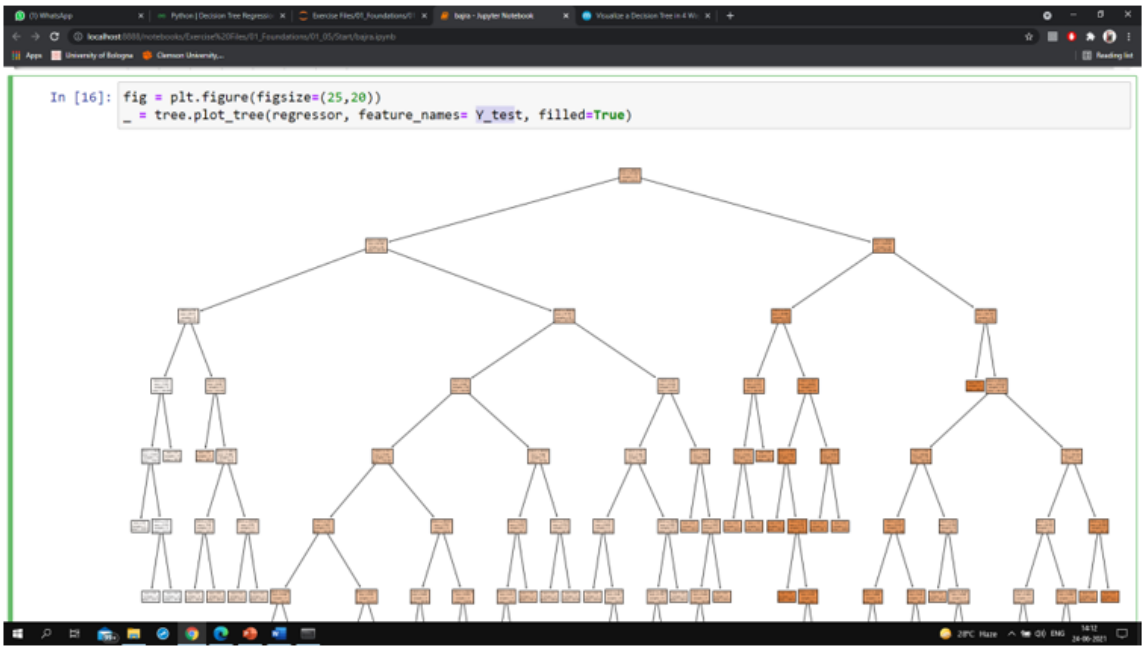

Fig. 8.2. Decision Tree representation

Decision Tree Regression Algorithm for Price Prediction

Price Prediction Decision Tree Regression Algorithm Decision tree regression is a machine learning regression technique that uses observation of an object's attributes and training of a model with a tree-like structure to forecast data in the future and generate useful continuous output. The output is not a discrete, well-known set of numbers or values when it is continuous. The algorithm's input is:

- The input parameter(current rainfall)
- The training dataset

➢ Formulas used for prediction

$$SSE = \sum_{i \in s1} (yi - y1) + \sum_{i \in s2} (yi - y2)$$

Where y1 and y2 are the values of the dependent variable in groups s1 and s2, which is a dataset parameter called the wholesale price index.Rainfall is the s1 and s2 group's equivalent.

The predictor values will be divided into groups in a recursive manner. When the split group's sample size reaches a particular point, the process comes to an end.

By substituting information gain with conventional reduction, the ID3 algorithm can be applied to the construction of decision trees for regression.

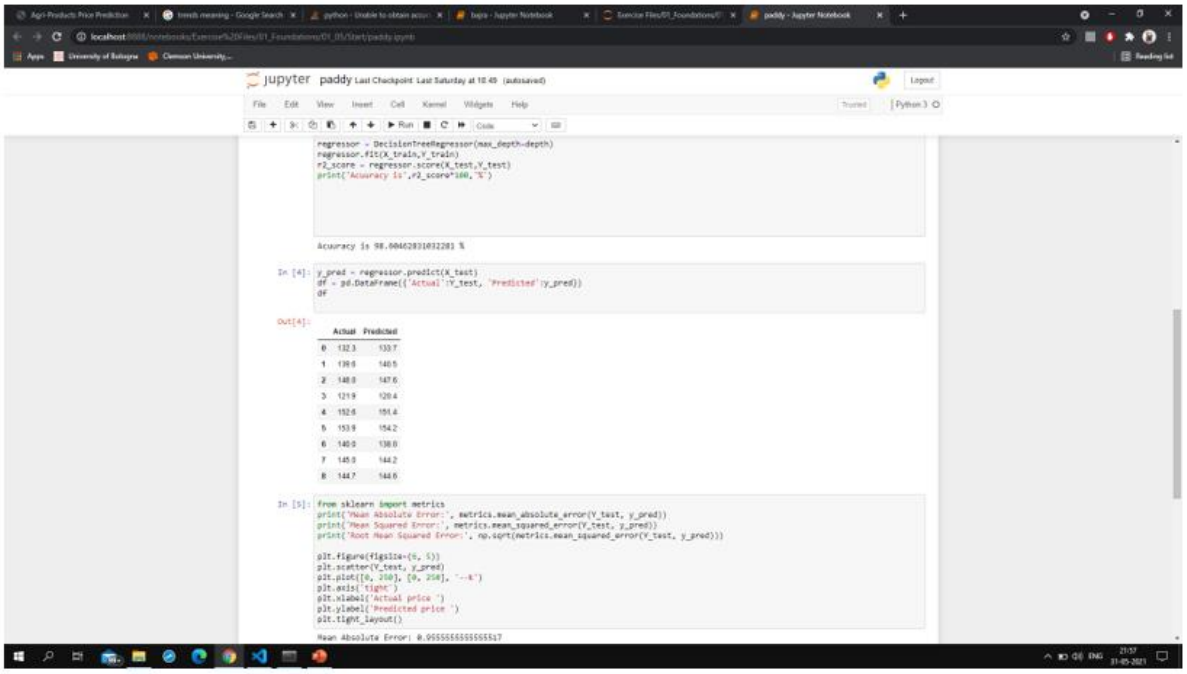

Fig. 9. Implementing Regression

STEPS TO IMPLEMENT THE ALGORITHM

Step 1: Initialize the dataset including training data for the wholesale price index and rainfall.

Step 2: Choose "x" as the independent variable and select all of the rows and column 1 from the dataset.

Step 3: Choose "y" as the dependent variable and select all of the rows and column 2 from the dataset.

Step 4: Fit the dataset to the decision tree regressor.

Step 5: Predict the new value.

Step 6: Visualize the outcome and assess its accuracy.

Max depth is the default value for the parameter that determines the size of the tree. The max depth parameter is used to decrease the complexity, size, and memory requirements of the trees.

Three equations are used to gauge decision tree regression's effectiveness. They are:

a. **Mean absolute error:** The average of all absolute errors is known as the **Mean Absolute Error (MAE)**. The equation is:

$$\mathbf{MAE} = \frac{1}{n}\sum_{i=1}^{n} |x_i - x|$$

Where:

- n = the number of errors,
- Σ = summation symbol (which means "add them all up"),
- |xi – x| = the absolute errors.

b. **Mean squared error :** You may determine how closely a regression line resembles a set of points using the **mean squared error (MSE)**. By squaring the distances between the points and the regression line, it is able to do this (these distances are the "errors"). Any unfavourable signals must be eliminated using squaring. Significant differences are also given more weight. It's called the mean squared error since you're calculating the average of several errors. The MSE decreases as the forecast gets better.

c. **R mean square score:** It is the mean of all errors squared, expressed as a square root. The root mean square error (RMSE) is recognised as a useful all-purpose error metric for numerical forecasts. RMSE should only be used to compare prediction errors of various models or model configurations for a single variable, not between variables, as it is scale-dependent.

OUTPUT:

This is the main page of the application that displays the welcome message.

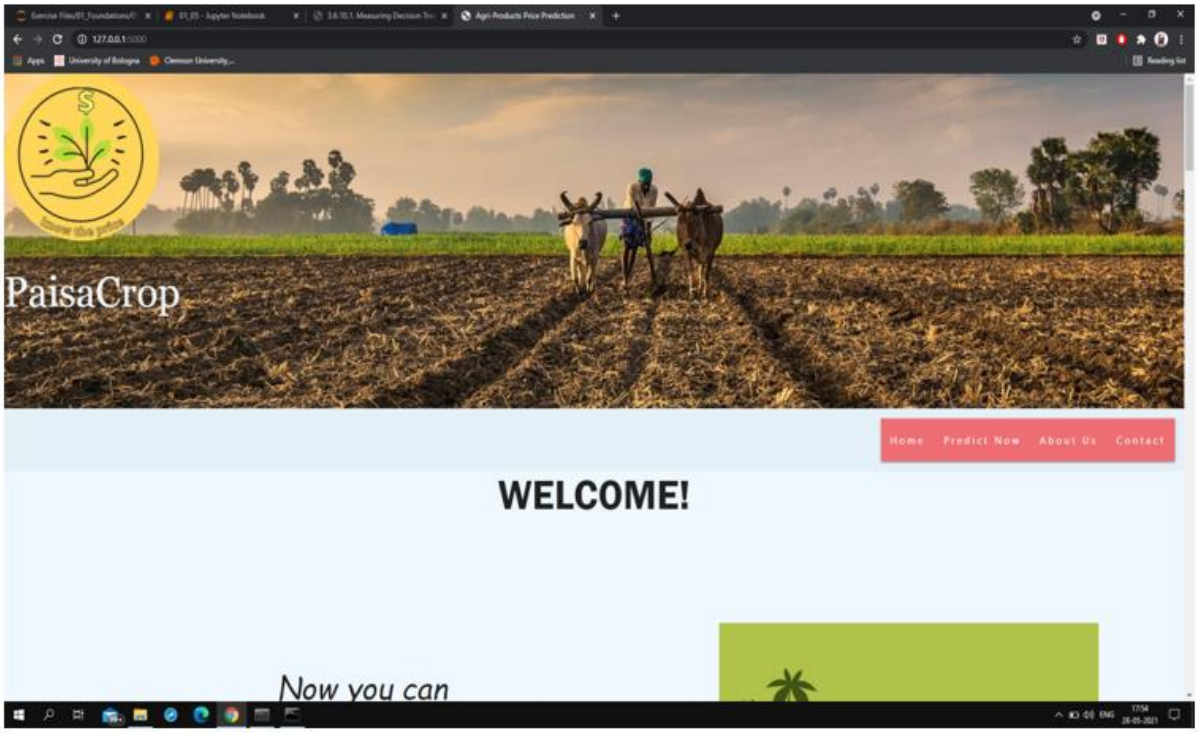

Fig. 10.1. Main Page

Displays the home page of the application.

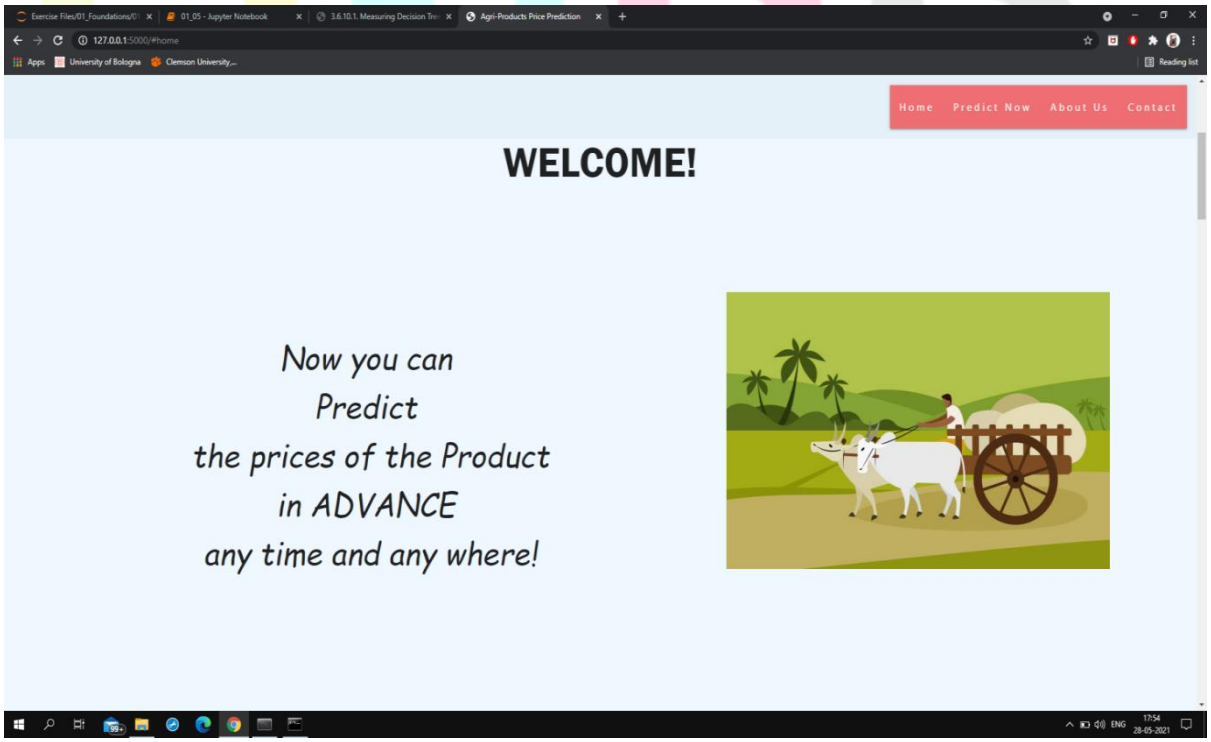

Fig. 10.2. Home page

This is the page where one can select the crops in order to predict the price.

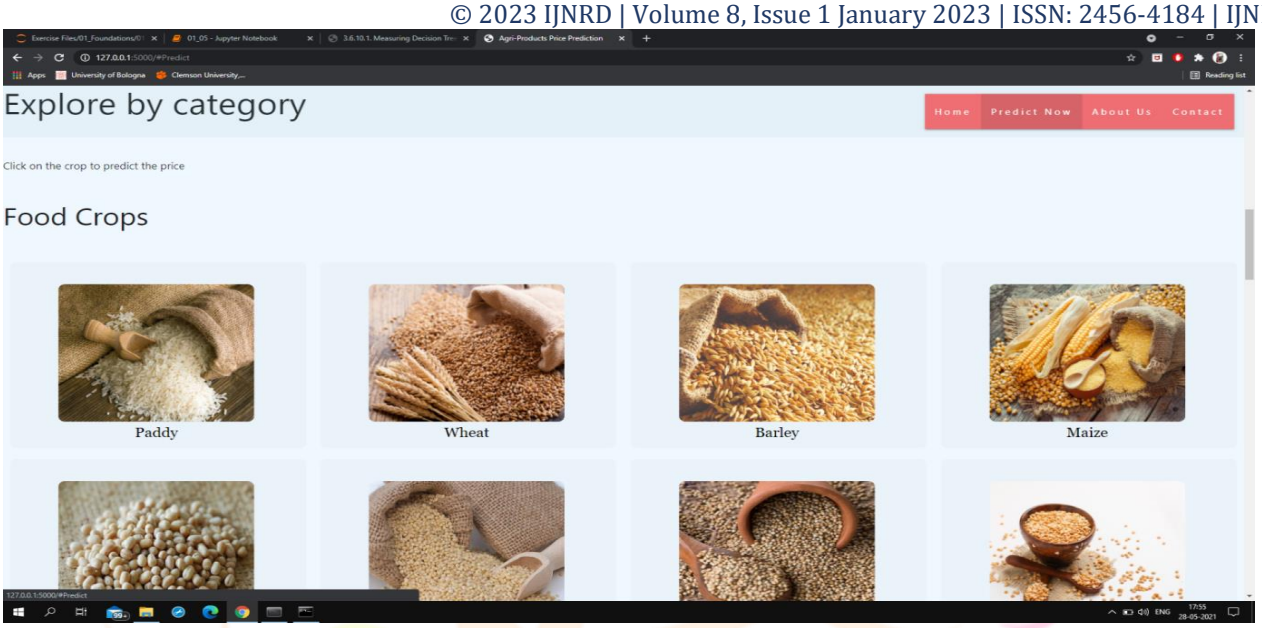

Fig. 10.3. Explore page

This page is used to select the Food Crops.

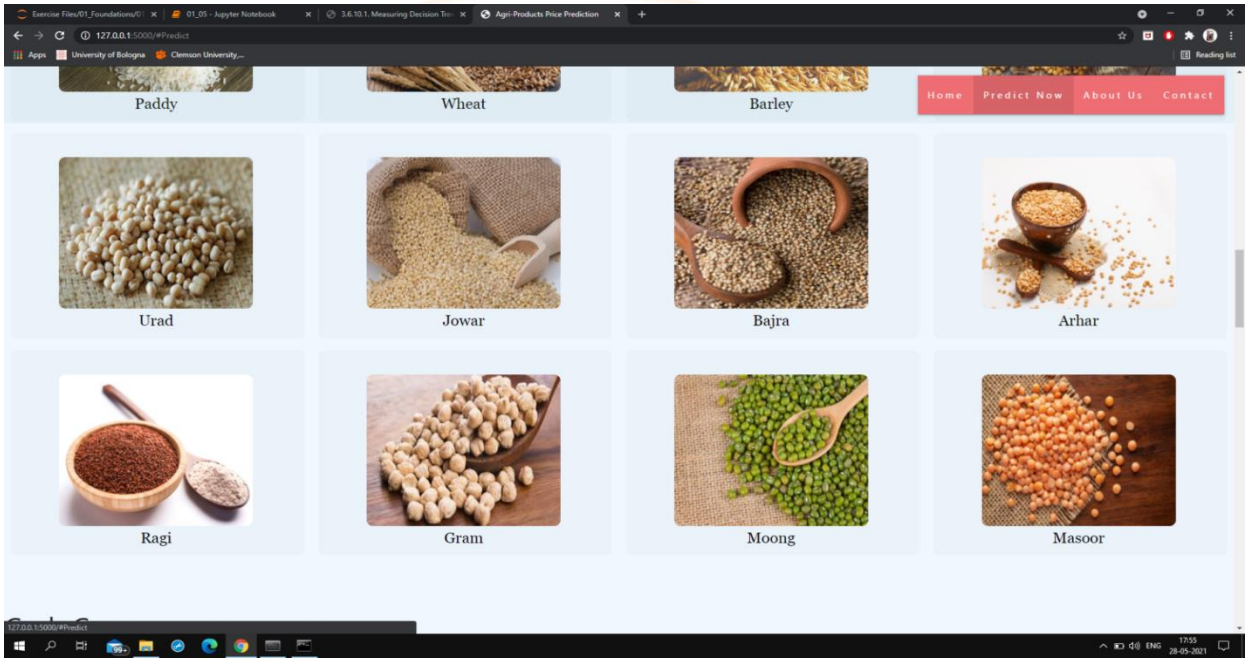

Fig. 10.4. Explore page -Food Crops

This page is used to select the Cash Crops.

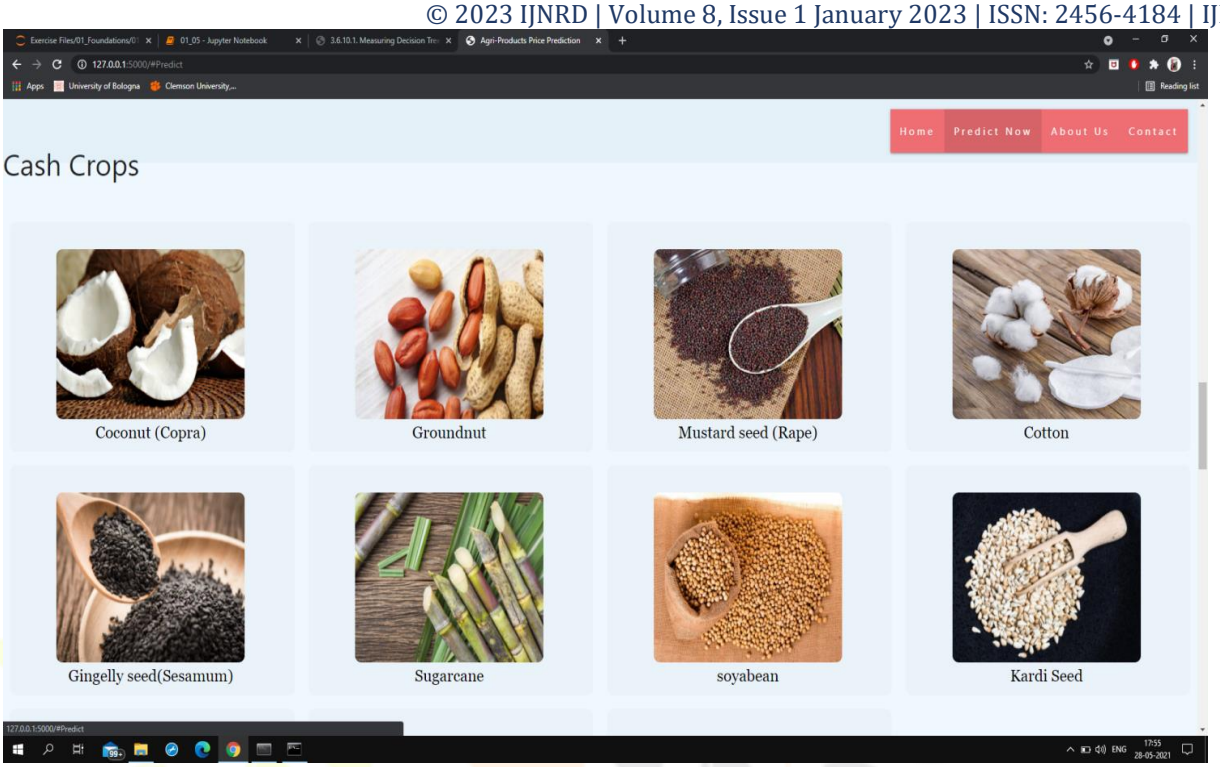

Fig. 10.5. Explore page- Cash crops

This page is the continuation of the Cash crops.

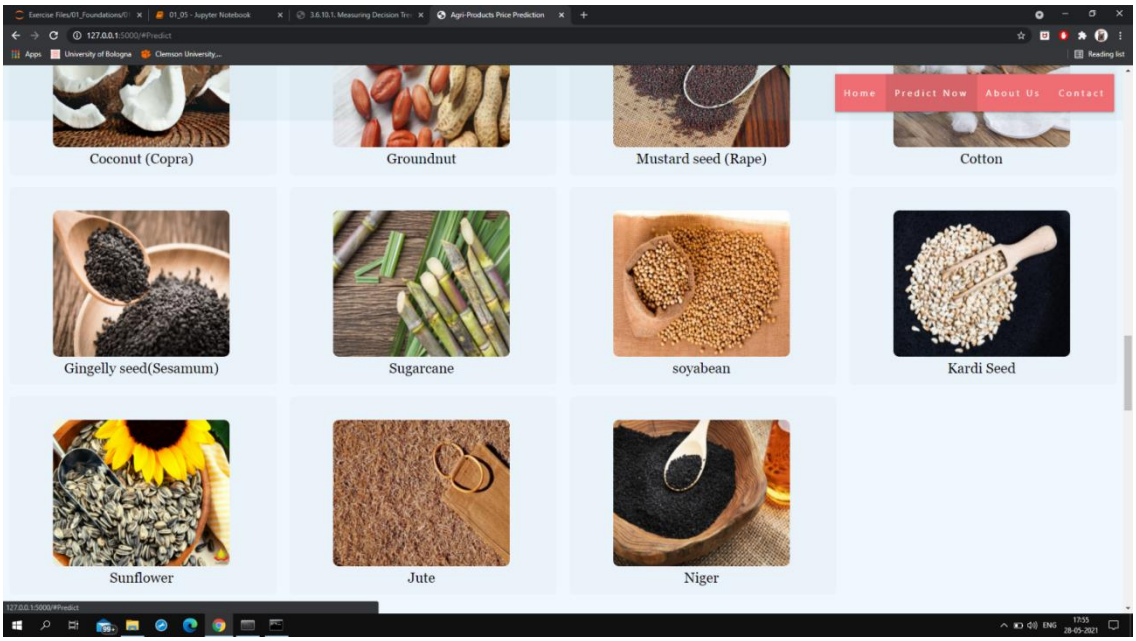

Fig. 10.6. Explore page – Cash Crops (continued)

This page displays the details about the application, like its vision and activities and also provides details about the current trending.

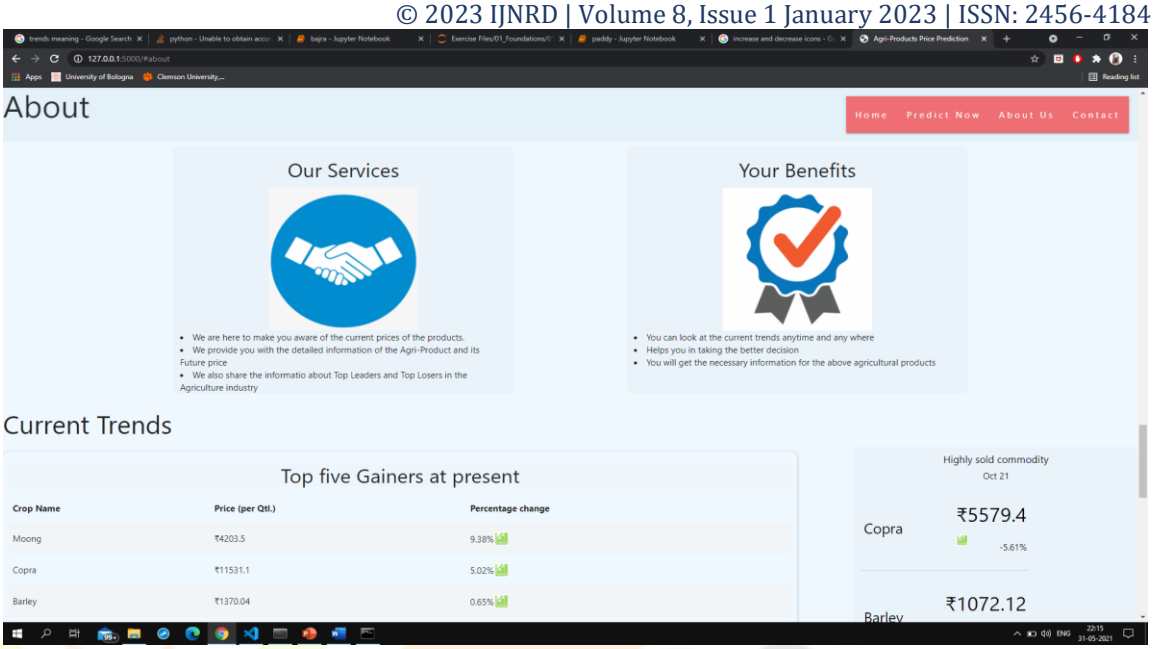

Fig. 10.7. About us page

This page displays the contact information of the creators or the servicers.

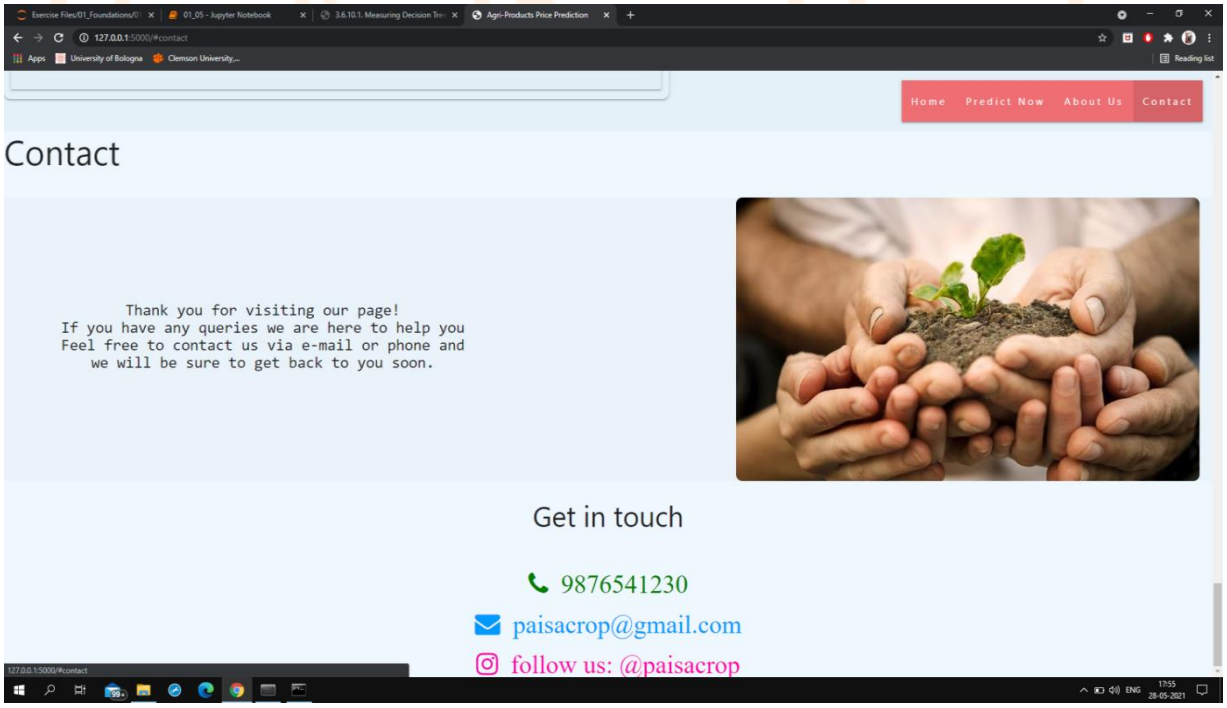

Fig. 10.8 Contact page

This is the result page after selecting the crop from predict page. It displays the future price of the product.

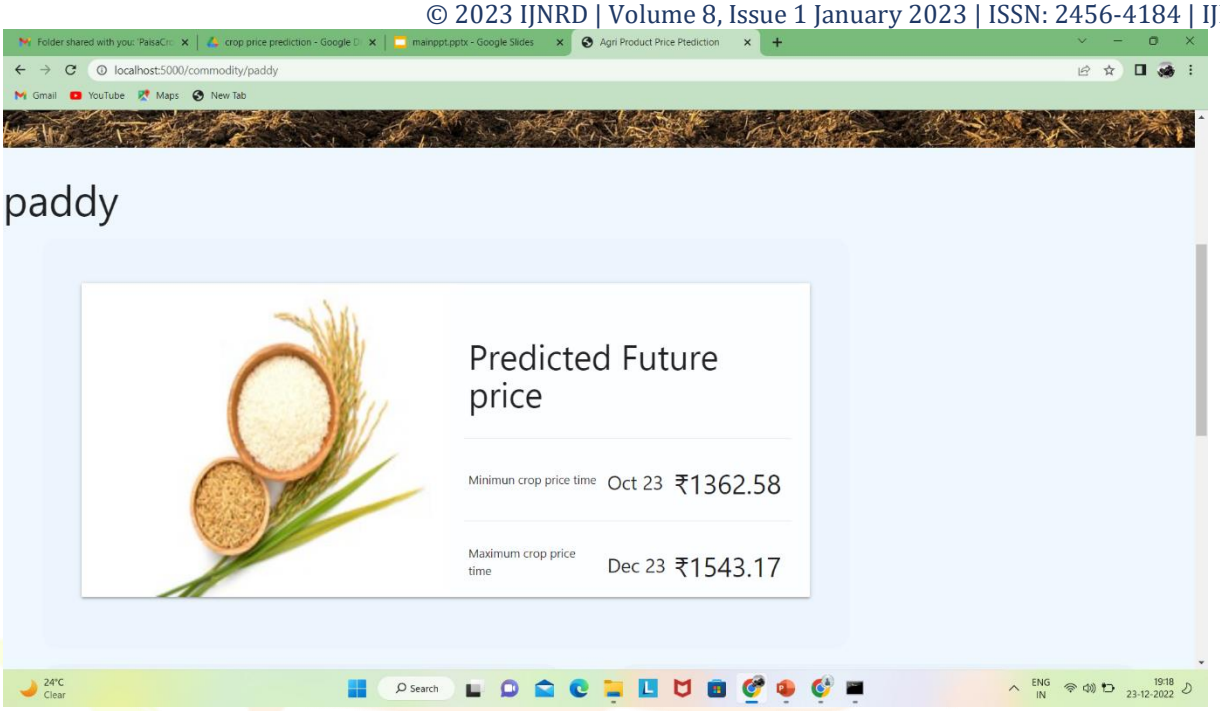

Fig. 10.9. Paddy result page

This is the result page after selecting the crop from predict page. It displays the current of the product along with the next twelve months prices.

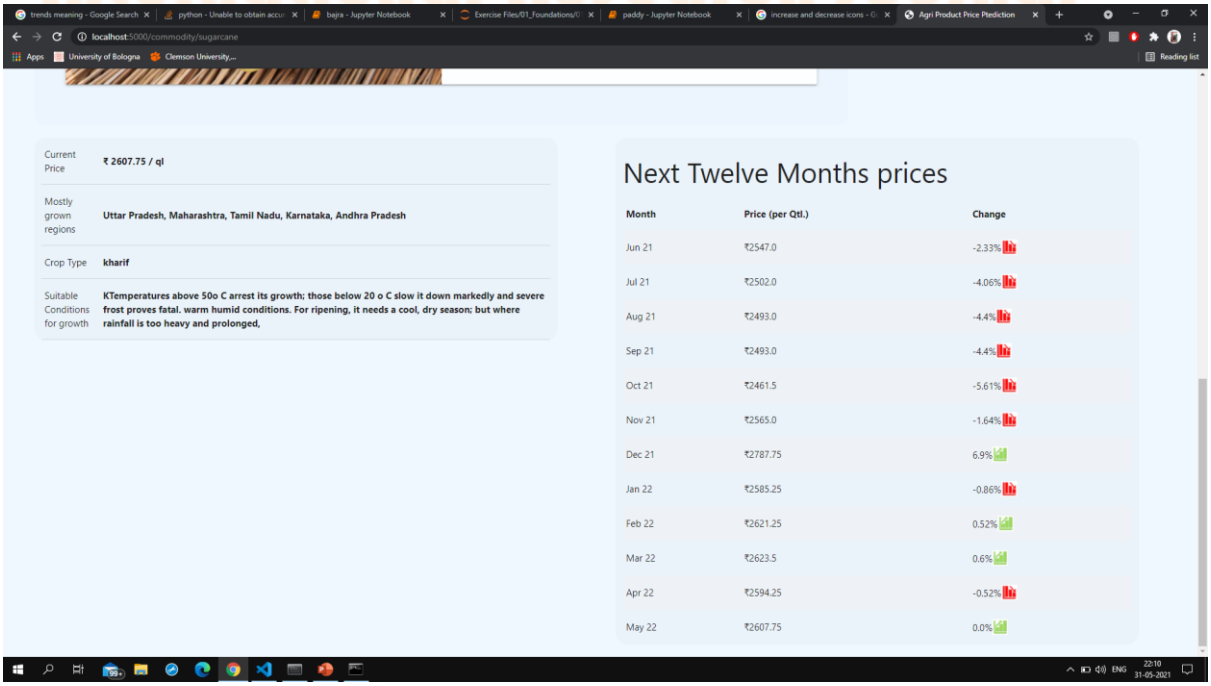

Fig. 10.10. Paddy result (continued)

TESTING

Testing is the process of assessing a system or its component(s) in order to determine whether it complies with the required specifications.

Additionally, it looks for flaws, gaps, missing features, or other requirements for a particular system. We have used the approaches of Integration testing and Unit testing for our project to test its functionality. In the instance

of unit testing, we examined each individual module to see whether or not the component pieces were accurate. Each module of the internal code has been tested. In the instance of integration testing, we tested the code for each module, combined them according to the hierarchy from top to bottom, and then tested the entire code to ensure that it passed each test case. We found a few errors in the final code during integration testing, fixed them, and retested the code numerous times until it passed all test cases.

Types of Testing

Testing is broadly classified in to two different types:

- Functional Testing
- Non Functional Testing

Functional Testing:

It is a kind of software testing that verifies the software system in comparison to the functional specifications or requirements. Each function of the software programme is tested using functional tests, which involve supplying the right input and comparing the output to the functional requirements.

Some of the Functional Testing are as follows

- Unit Testing
- Smoke Testing
- Sanity Testing
- Integration Testing
- White box testing
- Black Box testing
- User Acceptance testing
- Regression Testing

Non Functional Testing:

It is described as a type of software testing to examine a software application's non-functional features (performance, usability, dependability, etc.). It is intended to test a system's readiness according to non-functional criteria that functional testing never takes into account.

Some of the Non Functional Testing are as follows

- Performance Testing

- Load Testing
- Failover Testing
- Compatibility Testing
- Usability Testing
- Stress Testing
- Maintainability Testing
- Security Testing
- Disaster Recovery Testing
- Compliance Testing
- Portability Testing
- Efficiency Testing
- Reliability Testing
- Baseline Testing
- Endurance Testing
- Documentation Testing
- Recovery Testing
- Internationalization Testing
- Localization Testing

List of Test Cases:

S.NO	Test Name	Input	Expected output	Actual Output	Test Case Result
1	Loading the dataset	Dataset	Dataset Loaded	Dataset read	PASS
2	Split dataset	Training and testing dataset	Splitting of the data into training and testing	Dividing data into training and testing	PASS
3	Training the model	Training set, parameters	Trained with the provided set	Trained model	PASS
4	Validation of the model	No of entries from testing data	Validation of the model with best fit	Model generated	PASS

5	Predicting the accuracy	Accuracy metrics	Predicted accuracy	Accuracy Predicted	PASS
6	Test Data	Test Column in data	Prediction for the given test case	Predicted result	PASS
7	Test Case with invalid input	Invalid input values	Enter valid data	Please enter valid data	PASS
8	Starting with the Home page	Navigating to the home page	Redirecting to the home page	Redirected to home page	PASS
9	Starting up with the agri- product price Prediction Page	Choosing the product price prediction on the navigation bar	Redirecting to the agri-product price prediction Page	Redirected to the agri-product price prediction Page	PASS
10	Prediction of the agri -product price	Input values	Resulting Prediction	Prediction Provided	PASS

Table 1. Test cases of Application

CONCLUSION AND FUTURE SCOPE

Conclusions:

The primary goal is to foresee crop prices and, indirectly, to calculate the profit for the specified crops before planting. The obtained training datasets offer sufficient information for forecasting the proper price and demand in the markets. Following the completion of our project, we determined that the Decision Tree algorithm was the greatest fit because it provided the system with the maximum accuracy. When taking into account the precision of all the crops, the accuracy we attained is higher than that of the current system and is generally between 90% and 95%. The initiative has aided in the successful completion of the goal. As a result, the approach aids farmers in lessening their troubles and assisting them in making better decisions prior to selling their goods in the market, aiding them in preventing loss.

Scope for Future enhancements:

In order to further improve accuracy and probability, we want to instal our application in as many locations as we can (by using a GPS module) and extract the dataset specific to those locations. Another area for improvement is to implement chat portals to make the programme more user-friendly. Making sure that information is readily available to them and improving forecast accuracy raises the likelihood that farmers will make a profit.

REFERENCES

- https://www.mdpi.com/1424-8220/18/8/2674
- http://www.ieomsociety.org/singapore2021/papers/532.pdf
- https://wseas.com/journals/bae/2021/b865107-1274.pdf
- https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0270553
- https://irjiet.com/common_src/article_file/1593840444_849b1de2a7_4_irjiet.pdf
- https://www.academia.edu/32550508/Prediction_of_Future_Market_Price_for_Agricultural_Commodities_pdf
- https://ieeexplore.ieee.org/document/9047357
- https://www.emerald.com/insight/content/doi/10.1108/DTA-02-2021-0037/full/html
- http://www.aiggpa.mp.gov.in/uploads/project/Crop_price_predictions_using_machine_learning_in_MP-_A_pilot_study.pdf
- https://arxiv.org/ftp/arxiv/papers/2106/2106.12747.pdf
- https://www.ijitee.org/wp-content/uploads/papers/v9i2S/B12261292S19.pdf
- https://www.astesj.com/v06/i04/p42/
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9258887/
- https://www.kaggle.com/code/gcdatkin/vegetable-price-prediction/notebook
- https://www.journalijar.com/article/7567/forecasting-vegetable-price-using-time-series-data/
- https://iopscience.iop.org/article/10.1088/1755-1315/185/1/012013/pdf
- https://www.pantechelearning.com/product/agricultural-price-prediction-using-machine-learning/