Kisan Dhan - Crop Price Prediction Using Random Forest

Dr. Sonali Antad¹, Dr. Vipul Bag², Onkar Waghmode³, Shripad Wattamwar⁴, Atharva Wagh⁵, Aditya Zite⁶

¹,³,⁴,⁵,⁶Department of Computer Engineering
Vishwakarma Institute of Technology, Pune, 411037, Maharashtra, India
²Professor, N. K. Orchid College of Engineering & Technology, Solapur, Maharashtra, India

Article History:
Received: 18-04-2024
Revised: 30-05-2024
Accepted: 06-06-2024

Abstract:
Accurate prediction of agricultural commodity prices holds an important role for ensuring food security, profitability for farmers in farming, and making well-informed decisions for both farmers and industry stakeholders. Most of the prediction is made for farmers in the proposed system. The proposed system aims to find the relationship between weather conditions and agricultural prices by utilizing a comprehensive dataset spanning past years, including historical price data, modal, maximum and minimum prices, productivity, production and key meteorological things affecting like rainfall and temperature. The system also uses machine learning algorithms to classify the effects of climate factors, on price variations in combination with data collection. In addition to showing superior prediction capacity of the Random Forests than Decision Trees, this project is very good and major in terms of agriculture prices. These findings offer a good prediction for farmers in the agricultural industry to make secured decisions and face the challenges of price volatility. In a world where the stability of food production and economic sustainability depends on price predictability, this project contributes a practical and powerful tool for enhancing the prediction of the agricultural sector.

Keywords: Agriculture, Price trend, Decision Tree, Random Forest, Prediction.

1. Introduction

The agricultural sector has an important role to play in global food production and economic stability, as well as food security. In this ever evolving sector, it’s important to accurately anticipate and comprehend fluctuations in agricultural commodity prices. If the prediction is accurate then the farmers will significantly benefit, alongside with the advantages to the insurance companies and supply chain managers, to handle the supply chain management in a better way. Agricultural prices depend on multiple factors like market demand, geopolitical events, government policies and meteorological conditions like storms. The yields, quality and quantity of the crop shall be affected rapidly by weather factors such as rainfall and temperature. Therefore, analysis of these factors is essential for learning and anticipating price crises in the agricultural sector.

This proposed system focuses to explore agricultural price prediction by leveraging a rich dataset that covers over five years of historical price data, modal maximum and minimum prices, and critical meteorological variables. In order to enhance the accuracy and reliability of agricultural prices forecasts, different machine learning algorithms in particular Decision Trees and Random Forest are examined, with a view of identifying those that are best suited for accurately predicting farm prices.
In addition to the detailed analysis of how project works, the paper also includes advantages of proposed system over the existing systems. The website has been designed where the user can choose a crop, whose price has to be predicted.

2. Literature Review

This literature review includes some of the best research papers based on crop price prediction using various machine learning algorithms.

[1] Mainly in this research paper they have used 2 types of machine learning algorithm while doing actual price prediction. Those are decision tree and random forest algorithm. Also, they have used rainfall and Wholesale Price Index (WPI) data of last few years which makes their prediction more accurate. Their main goal is to identify price fluctuations and try to find out how prediction plays an important role in agriculture area.

[2] The focus shifts to predicting crop prices based on historical rainfall data and the Wholesale Price Index (WPI). In this paper, they have used Decision Tree Regressor, a machine learning algorithm, so with the help of this algorithm they are analyzing past data and making price predictions for the next 12 months. Their study shows that external factors such as rain and economic policy, how they play important roles in price prediction of crop in agriculture sector.

[3] In their project, machine learning models have played important role while predicting agricultural crop price, for reducing price risk for farmers. They achieved high accuracy rates, they achieved range of 87% to 97%, across multiple crops in Madhya Pradesh. Also, they provide their study to government of MP. Good forecasting through traditional and digital channels is important to increase the impact of these predictions and they actually provide help to farmers.

[4] This study focus the impact of price variation on farmers, also what is the need for price prediction models to decrease risk management within the agriculture supply chain. In terms of addressing these issues, the use of predictive analytics is promising in particular due to challenges such as unstable climatic conditions and global warming. The research focuses on the Areca nut crop in Kerala, India, and employs time-series and machine learning models such as SARIMA, Holt-Winter's Seasonal method, and LSTM neural networks to predict monthly prices. Among these models, the LSTM neural network exhibited superior performance based on the RMSE evaluation, offering a potential solution to help farmers navigate price volatility and make informed decisions.

[5] To predict the price of crops, how many crops are harvested in the past few years i.e. crop yield data and old price data are important factors. They have trained Random Forest and Linear Regression algorithms on the data they have. Their goal is to predict prices using data mining techniques and machine learning algorithms over the next few months. Previous year’s data plays a very important role in this project.

[6] This project is mainly designed to help mitigate farmer’s problems as a result of fluctuations in crop prices. The main goal is to make price prediction using Random Forest and Linear Regression algorithm. This algorithm makes predictions by analyzing historical data.

These projects demonstrate the growing interest in utilizing machine learning and data-driven techniques to enhance crop price-prediction models. They showcase the diversity of parameters and
datasets that can be leveraged to improve prediction accuracy. Furthermore, these studies underscore the potential benefits of such predictive models for stakeholders in the agricultural sector, aiding in decision-making and risk management.

3. Methodology

3.1 Proposed System

The proposed system offers a user friendly online interface for data entry as well as stringent preprocessing in order to ensure that the approach of crop price forecasting is not incompatible with each other. A machine learning model, which uses cutting edge algorithms for precise forecasts of crop prices on the basis of user inputs, is an essential component of this system and was developed using historical and weather forecast data. The proposed system also aims to provide region specific forecasts that focus on local market and environmental factors. Crop prices has to go above and beyond what actually price is given to by business people farmers and other agricultural stakeholders decision support and the information so they make risk ware decisions. In selecting the Random Forest algorithm over alternatives like Linear Regression and Decision Trees, several factors were considered. It is important to point out here that, for example, WPI, Metrological data, and previous year prices etc. The proposed system is essentially a machine learning model integrated with a web application. It primarily addresses the problem of crop price variability and attempts to address it by providing crop price predictions prior to crop sowing. The model is trained using datasets provided by the Department of Agriculture and Cooperation of India and the Office of Economic Advisor, Government of India. It is based on the random forest algorithm. As a result, the proposed system offers the user-specified commodity's trend in crop prices as well as the prediction of crop prices in specific months and years. This envisioned crop price system is specifically designed with a focus on Indian crops and localities.

3.2 System Architecture of Proposed System

![System Architecture of Proposed System](image)

Fig. 1. System architecture of proposed system consist of different stages shown in above figure.
This is a system architecture of proposed system. It designed to predict crop price forecasts according to commodity, month, year and rainfall. Information regarding the price variable such as natural conditions, Government policies, production of commodity and seasonal temperature is made accessible to users in a user friendly online portal of the envisaged system. Following this vital preprocessing stage, the user-provided raw data is carefully scaled to ensure compliance with proposed machine-learning model. Proposed system’s prediction engine is a model that was developed using previous data. Based on the supplied input features, it makes precise crop price predictions using well known ml algorithms such as random forest and decision tree algorithm. When forecasts are made, the outcomes are smoothly returned to the web application so that users may get insightful information on crop price patterns. Proposed method offers stakeholders in the agriculture industry a potent tool for well-informed decision-making and risk management by utilizing the power of machine learning for real-time crop price prediction.

3.3 Technical Requirement

Machine Learning, Data Science and Web Technology have been incorporated into the proposed system. Some libraries are being used for the building of this system, which include NumPy, Pandas, Scikit-learn, Matplotlib. The arithmetic and statistic operations in the system are being made by NumPy. It supports the arithmetic and statistical and logical operations on large multi-dimensional arrays. Pandas is used for analyzing, cleaning, exploring, and manipulating data basically pandas library is used for data preprocessing. Scikit-learn includes many regression algorithms such as linear regression, Decision tree, Random forest and many other algorithms. The main use of these algorithms are analyzing the dataset and finding some patterns and insights from data and according to this information predicting the target value. Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Basically Matplotlib is used to analyze the relationship of attributes using graphical way and also it is used for testing the model by plotting the
The proposed crop price forecast technology goes through a series of steps to ensure that it works correctly. The process involves data gathering, data cleaning, splitting for training and tests, algorithm evaluation, and thorough validation of accuracy prior to integrating model into the web application.

3.4 Workflow of Proposed System

The proposed crop price forecast technology goes through a series of steps to ensure that it works correctly. The process involves data gathering, data cleaning, splitting for training and tests, algorithm evaluation, and thorough validation of accuracy prior to integrating model into the web application.

Fig. 3. System design and development were done in the order mentioned above.

a. **Data collection** – Data is very crucial component in the field of machine learning. According to the dataset machine learn the patterns and insights and produce the result so accurate dataset is very important to build robust and effective machine learning model. For accurate prediction data for proposed system is collected from authorized government websites such as data.gov.in and from APMC and local mandis. This data set include following attributes I) WPI Index II) rainfall (in mm) III) year and month this data collected range includes data from year 2012 to year 2018.

b. **Data Preprocessing** – Data preprocessing is very important step in machine learning that includes Data Cleaning, Data encoding, Data scaling and Data transformation. Data preprocessing is required because some specified machine learning algorithms need data in special format such as Random forest algorithm does not support null values in dataset so to execute random forest algorithm null values have to be managed from original raw dataset. Pandas library is very helpful while dealing with missing values. Data cleaning mainly includes removing of unwanted observations, managing unwanted outliers and fixing structural errors. Data cleaning improve the effectiveness and perfection of dataset so that machine
learning model train on accurate dataset in result gives accurate and precise results. Data encoding is another important factor in data preprocessing it converts the categorical values into numeric format. Many machine learning model only work on numeric values so data encoding is must step in data preprocessing pipeline. Basically Data preprocessing converts the original raw data into the data that capable of being processed using various machine learning models and algorithms.

c. **Model Testing** – After data processing next important step is to train the dataset using various machine learning algorithms and finding the best performer algorithm that gives accurate and precise results. For training the model dataset divided into train and test dataset with 8: 2 ratio. After splitting the dataset model training is very crucial part in machine learning in model training It must be ensured that the system is capable of coming up with accurate forecasts before implementing it in practice. To find the best algorithm for making the forecast as accurate as possible, a number of approaches and algorithms were being tested. Unlike Linear Regression, which assumes linear relationships between variables, and Decision Trees, which can be prone to over fitting, Random Forest excels at capturing intricate and nonlinear patterns within the data. Random Forest works by constructing a lot of decision trees and aggregating their predictions. This method followed in less variation and improving prediction stability. The versatility of Random Forest in managing a wide variety of input factors and complicated relationships within the data is very useful in the context of crop price prediction. Below table shows the errors of various algorithms after implementing them.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Root Mean Square Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Regression</td>
<td>4.675</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>3.452</td>
</tr>
<tr>
<td>Random Forest</td>
<td>1.674</td>
</tr>
</tbody>
</table>

Root mean square error is most commonly used measures for the evaluating the quality of the predictions.it shows the difference between the measured true value and predicted value using Euclidian distance. As above table shows random forest algorithm gives minimum error so maximum accuracy so this system model is trained using Random forest algorithm and approximately random forest algorithm gives 95 – 98% accuracy so proposed system gives proper results with 95-98% accuracy. Here proposed system predict the WPI index and according to WPI index system calculate the price of specified commodity according to following equation -

Predicted price = (WPI Index * Modal price of crop)/100.

Here model price of crop is considered as the MSP (minimum support price) of the crop that is decided by the department of Agriculture and co-operation, Government of India. This is how the system forecasts the prediction price of commodity. This extensive procedure guarantees that the proposed crop price prediction tool is helpful to those working in agriculture, letting them make judgments.
regarding crop prices and market trends that are both trustworthy and beneficial. While training the model all the factors are considered that directly gives impact on crop price such as natural conditions, rainfall, temperature, humidity most important is period or seasons such as Rabi and Kharip and proposed system also focuses on government policies that is also impactful factor. As considering all the factors and conditions model train accurately and gives very accurate and precise predictions.

4. Results and Discussions

The Proposed system is the crop price prediction system that predict the commodity price and predict the trend of price in future and guide the farmers into right direction so that they can gain more profit from the crop. As India is the Agricultural country and India has the large agriculture sector and most of the houses in India economically depends on the agriculture. Total 70% of population of India depends on the agriculture for their livelihood. India is known as the country of farmers, despite the fact that the number of Indian farmers dying on a daily basis is increasing. The most recent statistics available show that nearly thirty farmers commit suicide every day, usually as a result of overwhelming debt. In fact, government data from 2020 shows that over 10,000 farmers took their own lives. The main reason for suicide of the farmer is economic crises. The main goal of this suggested system is to give farmers financial security by supplying crop price forecasts for every commodity, enabling them to choose which crops to sow and which not to. Farmers benefit from the system in two ways: first, they earn from the crop's production; second, they are protected against loss because the system provides the correct guidance in the first place. This is an efficient strategy to lower the rate of farmer suicide in India and contribute to the development of a stronger and more stable Indian economy. Farmers can also increase their profits and achieve economic stability in this way.

System GUI of predicted result –

Fig. 4. User Interface of display of predicted price

Fig. 5. Price forecasting of upcoming year.

Fig. 6. Price forecasting of current year.

Fig. 4 is the user interface of result displayed to user here system not only gives the predicted price but also provides the Average, minimum and maximum price of specified commodity according to various
mandis here system also provides the Maximum price and minimum price of commodity in overall year with month. Fig 5 display the graph of price forecast given by system it shows the trend of the price of commodity in upcoming year to user and in Fig 6 display the graph generated by system that provides the trend in crop prices in current year. This is how system provides the overall information of crop price from current year to next year.

4.1 Experimental Results -

![Scatter Plot](https://internationalpubls.com)

**Fig. 7. Scatter plot between labeled (original) and predicted value graph**

Above graph represent the result of random forest algorithm in the form of graph between original and predicted WPI Index. It shows the linear graph between predicted value and actual value that means most of the predicted values matches with the actual values that means prediction accuracy of system is very high and also there are not any outliers in the graph most of the points follow the linear curve so precision of prediction also goes well. Below table shows the statistics of prediction with different years.

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual Value</th>
<th>Predicted Value</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>98.2</td>
<td>98.188</td>
<td>~ 0.01</td>
</tr>
<tr>
<td>2013</td>
<td>98.4</td>
<td>98.006</td>
<td>~ 0.2</td>
</tr>
<tr>
<td>2014</td>
<td>112.8</td>
<td>113.4</td>
<td>~ 0.5</td>
</tr>
<tr>
<td>2015</td>
<td>115.7</td>
<td>114.036</td>
<td>~ 1.66</td>
</tr>
<tr>
<td>2016</td>
<td>120.1</td>
<td>120.76</td>
<td>~ -0.06</td>
</tr>
<tr>
<td>2017</td>
<td>126.1</td>
<td>125.199</td>
<td>~ 1.00</td>
</tr>
<tr>
<td>2018</td>
<td>118.3</td>
<td>119.49</td>
<td>~ -1.1</td>
</tr>
</tbody>
</table>
Above table represent the actual and predicted WPI index with difference between them here table represent the system prediction accuracy with respect to various years. It shows that the stability in the accuracy of system it shows that accuracy of prediction maintain the stable behavior with successive years. As Predicted WPI index is accurate so crop price determined from it also very accurate and very close to actual crop price.

Below graph represent the predicted price trend of Jowar in the year of 2023 with average and maximum price of Jowar in each month.

The suggested system aids farmers in choosing when to plant crops by offering a crop price prediction trend for the entire year. Here developed system also able to predict the price trend of upcoming years below is the example price trend of Jowar crop in the year 2024 –

Here above graph represent the Jowar crop price trend in year 2024. The Jowar crop price trend shown in the above graph corresponds to the year 2024. The system in this case employs rainfall values anticipated from reputable websites, which are the primary determinant. As a result, the rainfall data
utilized must be extremely accurate to have a direct impact on crop price prediction.

4.2 Crop price analysis of Jowar crop with respects to seasons

Rabi Season - Jowar cultivation commences during the months of June to July. The harvesting of Jowar typically spans over a 5-month period, culminating in December. Given that December is the designated harvesting month, it follows that Jowar prices tend to remain stable in the subsequent months, namely January, February, March, and April.

Kharif Season - Jowar sowing for the Kharif season takes place in September to October. The harvesting of Jowar in this season also extends over 5 months, concluding in April. Consequently, Jowar prices are observed to remain stable in the months of May, June, July, and August. Considering both the Rabi and Kharif seasons, it can be inferred that the Jowar prices exhibit stability during the months of January through August. Moreover, after August, the rainfall tends to decrease.

Fig.10. WPI index by government of India for 2023 published by office of economic adviser

The WPI (Wholesale Price Index) exhibits a stable trend from January to August, after which it gradually begins to show an upward trajectory. This serves as an evidence for the accuracy of the predicted values. The most important factor in the crop price prediction is rainfall (mm) from which the price has been predicted. In order to make this forecasting system predict for 2024, the predictions have been made with the help of other websites that forecasts rainfall. As analysing the result of prediction with respect to the seasons it shows that the system follows the pattern that actual trend follows it also consider the impact of seasons on the crop prices.
4.3 Data-Driven Agricultural Decision Support System: Breaking the Cycle of Crop Overproduction

The proposed system offers a comprehensive display of the current distribution of crop sowing across various regions, providing farmers with valuable insights to make informed decisions about their own crop choices. The primary objective is to address the prevalent traditional mindset among farmers, wherein a high demand and price for a particular crop in one year lead to a widespread adoption of that crop in the following year. This collective decision results in an oversupply of that crop, causing a significant drop in prices. Consequently, farmers who heavily invested in that crop face substantial losses. To tackle this issue, the proposed system introduces a feature that visually represents the current proportion of sowed crops in different regions, offering farmers a clear view of trends in crop selection. By presenting graphical data on the predominant crops being sowed, both at the overall level and within specific regions, the system enables farmers to make well-informed decisions based on the choices of their peers. This data-driven approach allows farmers to assess the prevailing trends in crop sowing and adjust their choices accordingly, mitigating the risk of overproduction and subsequent financial losses. By facilitating a more informed decision-making process, the proposed system aims to break the cycle of blindly following high-
demand crops, thereby promoting a more balanced and sustainable agricultural landscape.

5. Future Scope

• Emerging technologies such as Artificial Intelligence (AI), Deep Learning, and Neural Networks can be incorporated to improve the accuracy and efficiency of prediction models.

• Reinforcement learning is also play an crucial role to update the dataset with current prices it helps to train the model with current data so that model can learn and find the new patterns from dataset that result into more accuracy.

• An evolution toward real-time data integration through Internet of Things (IoT) devices installed across agricultural landscapes, weather stations, and market channels will revolutionize prediction accuracy, ensuring dynamic factors like weather fluctuations and market trends are instantly factored into forecasts.

• A significant development will be the implementation of area-wise predictions. By considering the unique environmental and market conditions in specific geographic regions, the project will provide tailored forecasts.

6. Conclusion

The Crop Price Prediction using random forest project has named as “Kisan Dhan”. Kisan Dhan has been a valuable step in improving the predictability of agricultural commodity prices, which is essential for ensuring food security, sustainability on farms and riskless decision making by farmers and industry representatives to make sound decisions going forward, and to ensure the best results, it also plays a role in supply chain management. The main objective of this proposed system is to harness data-driven methodologies to forecast crop prices accurately, providing valuable insights and empowering stakeholders in the agricultural sector to make informed decisions for improved productivity, profitability, and sustainability. The employed machine learning models showcased notable accuracy in forecasting crop prices. The models were able to analyze historical data patterns and make reliable predictions for future prices of various crops. Use of in Random Forest machine learning algorithm resulted more accurate crop price prediction. Kisan Dhan emerges as a comprehensive solution to address the financial challenges faced by farmers, offering a crucial foundation for their economic stability. By providing insights into the price trends of commercial crops, Kisan Dhan essentially functions as a guide for crop selection and cultivation. This approach translates into increased profits for farmers, thereby contributing to the stabilization of their financial conditions. Given that approximately 70% of the Indian population relies on agriculture, the stability of farmers' conditions directly impacts the nation's economy. The rising rate of farmer suicides and a decreasing interest among the youth in pursuing agriculture as a career have further eroded the identity of India as an "agricultural country." Kisan Dhan aims to reverse this trend by introducing a tech-based farming approach that goes beyond traditional methods. The proposed system not only offers farmers a comprehensive overview of crop price trends but also empowers them to make informed decisions.
about which crops are more lucrative. This proactive approach allows farmers to plan their future endeavors based on market dynamics, ultimately leading to increased profitability. As the financial outlook for farmers improves, the youth of the country becomes more inclined to embrace agriculture as a viable career option. Kisan Dhan, by promoting modern and profitable farming practices, rekindles interest in agriculture among the youth. This shift not only revitalizes India's identity as an "agricultural country" but also strengthens the economic foundation of the nation. In essence, Kisan Dhan emerges as a transformative initiative, fostering a sustainable and prosperous future for both farmers and the nation as a whole.

7. Acknowledgement

Expressing gratitude to the data.gov.in website for providing essential datasets, appreciation is extended to all MANDIs and APMCs markets for their assistance in data collection for machine learning purposes. Thanks are conveyed to the Department of Agriculture and Cooperation, Government of India, for their contribution in sharing the WPI index data. Furthermore, acknowledgment is given to agricultural guides for their insightful information on crop price variations and associated factors.

References

[3] State agriculture board & abhioggapa’s “Crop price predictions using machine learning in MP_A_pilot_study” Project report. Published in State Agricultural Marketing (Mandi) Board, Gov of MP.
[8] India Meteorological Department., rainfall data of month wise in various states.


Pradhan, Durgesh & Kasera, Mayur. (2018). A REVIEW ON IMPACT OF TECHNOLOGY IN AGRICULTURE OF INDIA.

