IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125197

RAITHA SAATHI - An AI/ML-based application for market price and demand prediction

Dr. Archana B[1], Abhishek D Badiger[2],Rohit R Aradhya[3], Sonika KV[4], Varsha MG[5]

Computer Science and Engineering Vidya Vikas Institute of Engineering and Technology, Mysore[1]

Student of Computer Science and Engineering Vidya Vikas Institute of Engineering and Technology, Mysore[2-5]

Abstract: Accurate prediction of crop prices is crucial for farmers to make informed decisions about agricultural production and trade. It enables them to optimize their planting decisions, determine the harvest time, and plan their sales strategy to maximize profits. Given the increasing volatility of agricultural markets due to climate change and other factors, the significance of crop price prediction has grown in recent years. As a result, sophisticated models that can analyze large amounts of data to provide accurate price forecasts are in high demand. Nowadays, farmers seek to leverage analytics for obtaining the data they need to make actionable insights and informed decisions. Automated farming is becoming more popular among farmers in many countries. Crop price estimation and evaluation are critical to minimize losses and manage the risk of price fluctuations when farming a specific crop type. Falling prices can lead to significant losses for farmers. In this study, we employed the Random Forest Algorithm to analyze past data, predict prices for new data, and estimate crop prices.

1. INTRODUCTION:

The objective of agriculture is to generate profits while working within the constraints of limited land resources. Crop price prediction plays a vital role in this process by forecasting the future prices of agricultural commodities. Traditionally, farmers relied on their experience in a specific crop field to make predictions. However, due to changing conditions, it has become necessary for agricultural practices to progress and adapt to the evolving landscape. This entails analyzing historical data, current market conditions, and various economic and environmental factors that can influence crop yields and prices.

Accurate crop price prediction is of utmost importance for the effective functioning of agricultural markets and enables farmers to make well-informed decisions. By gaining insights into and predicting crop prices through the utilization of advanced techniques, productivity can be enhanced. In our approach, we employed the Random Forest Algorithm to analyze previous data, predict prices for the latest data, and estimate future prices.

2. LITERATURE SURVEY

Numerous strategies have been employed to enhance the financial outcomes of agricultural yields. Our system development involves a thorough analysis and consideration of several notable approaches.

[1] To gather the necessary data for this system, information on relevant crops is collected from local markets and through online surveys. This data is then used as a dataset to train Machine Learning Models. Price prediction is performed utilizing algorithms such as Artificial Neural Networks, Partial Least Squares, and Autoregressive Integrated Moving Averages. Based on the analysis of recent data, Partial Least Squares, and Artificial Neural Networks demonstrate superior performance in both short and long-term prediction compared to the aforementioned algorithms.

[2] The objective of this paper was to assist farmers in making informed decisions by providing a ranking of crop suitability for a specific area. Prediction and ranking are achieved through the utilization of supervised machine learning techniques, including the K-nearest neighbor regression algorithm and decision tree learning.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁘ Impact Factor 8.102 ⁘ Peer-reviewed / Refereed journal ⁘ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125197

3. RANDOM FOREST ALGORITHM

Random Forest is a popular ensemble machine-learning algorithm used for predictive modeling. It is part of the ensemble method family, which combines multiple models to improve predictive performance by reducing bias and variance.

By constructing multiple decision trees during training and outputting the mode of their classes, Random Forest provides a more stable and accurate prediction. To prevent overfitting specific features, the algorithm randomly selects a subset of features and builds decision trees on them. Bootstrapping generates multiple datasets from the original data, allowing each decision tree to be trained on a slightly different subset.

During prediction, the algorithm aggregates the predictions of all decision trees and calculates the importance of each feature by evaluating how much error would increase if feature values were randomly permuted. This helps identify the most critical features of the model.

Data Sampling: Random Forest randomly selects a subset of the original data with replacement (bootstrapping). Each sample is of the same size as the original dataset.

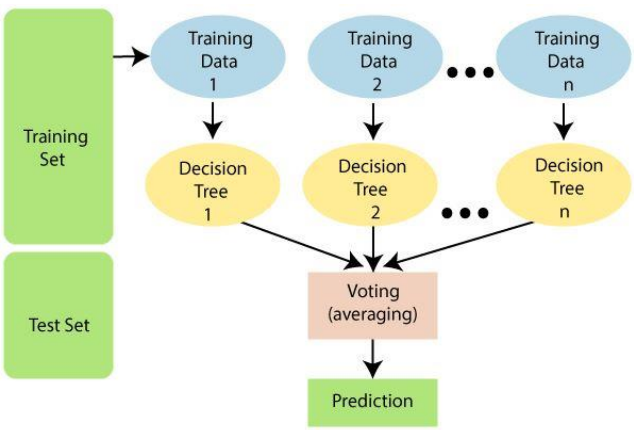
Feature Sampling: For each decision tree, a random subset of features is selected from the available features in the dataset. This helps to reduce the correlation between trees and improve the model's variance and accuracy.

Tree Building: For each decision tree, the algorithm recursively splits the data based on the selected features and their corresponding split points. The splits are made to maximize the reduction in impurity (e.g., entropy, Gini index) of the target variable. The tree continues to grow until a stopping criterion is met.

Prediction: During prediction, the algorithm aggregates the predictions of all decision trees by either taking the majority vote (classification) or the average (regression).

Feature Importance: The Random Forest algorithm can also calculate the importance of each feature in the dataset. This is done by computing how much the error of the model would increase if the values of a particular feature were randomly permuted. The higher the increase in error, the more important the feature is to the model.

Evaluation: The model's performance is evaluated using metrics such as accuracy, precision, recall, F1 score, mean squared error, etc. If the model's performance is not satisfactory, the hyperparameters of the algorithm can be tuned, and the process is repeated until the desired accuracy is achieved.



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125197

Fig Random Forest Algorithm

4. IMPLEMENTATION

Data Collection: data collection involves gathering information on various factors that influence crop prices, such as historical prices, weather conditions, soil quality, and economic indicators (refer to Fig 4.1).

	A	B	C	D	E	F	G	H	I	J
1	state_name	district_name	market_center_name	Variety	group_name	date_arrival	Month	Year	Arrival	MODAL
2	Karnataka	Mysore	T. Narasipura	Pachha Bale	Fruits	18 Jun 19	06	19	6	2880
3	Karnataka	Mysore	T. Narasipura	Pachha Bale	Fruits	11 Jun 19	06	19	10	29
4	Karnataka	Mysore	K.R.Nagar	Medium	Fruits	29 Aug 18	08	18	2500	13000
5	Karnataka	Mysore	K.R.Nagar	Medium	Fruits	17 Aug 18	08	18	1500	13000
6	Karnataka	Mysore	K.R.Nagar	Medium	Fruits	28 Jul 18	07	18	1500	13300
7	Karnataka	Mysore	Mysore (Bandipalya)	Pachha Bale	Fruits	6 Feb 23	02	23	3	1400
8	Karnataka	Mysore	Mysore (Bandipalya)	Pachha Bale	Fruits	27 Jan 23	01	23	3	1500
9	Karnataka	Mysore	Mysore (Bandipalya)	Pachha Bale	Fruits	26 Jan 23	01	23	9	1500
10	Karnataka	Mysore	Mysore (Bandipalya)	Elakki Bale	Fruits	21 Jan 23	01	23	6	1400
11	Karnataka	Mysore	Mysore (Bandipalya)	Elakki Bale	Fruits	20 Jan 23	01	23	4	1500
12	Karnataka	Mysore	Mysore (Bandipalya)	Pachha Bale	Fruits	18 Jan 23	01	23	15	2200
13	Karnataka	Mysore	Mysore (Bandipalya)	Elakki Bale	Fruits	17 Jan 23	01	23	15	1450
14	Karnataka	Mysore	Mysore (Bandipalya)	Pachha Bale	Fruits	16 Jan 23	01	23	15	1450
15	Karnataka	Mysore	Mysore (Bandipalya)	Pachha Bale	Fruits	6 Jan 23	01	23	7	1500
16	Karnataka	Mysore	Mysore (Bandipalya)	Pachha Bale	Fruits	21 Dec 22	12	22	19	1410

Fig 4.1

Data Pre-Processing: this includes removing any irrelevant or missing data, normalizing the data, and converting categorical variables into numerical ones.

Python programming language offers numerous libraries for data analysis and visualization, including Pandas, which facilitates efficient data manipulation and analysis. Additionally, NumPy provides support for large multi-dimensional arrays and matrices, coupled with an extensive collection of mathematical functions for working with these arrays. To create high-quality data visualizations that represent information effectively, users can leverage Matplotlib, a plotting library that integrates seamlessly with NumPy.

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestRegressor
```

```python
dataset = pd.read_excel('D:/Python programs/New Crop Prices.xlsx')
```

Fig 4.2 shows the different Python libraries used for the prediction.

```python
dataset.head()
```

	state_name	district_name	market_center_name	Variety	group_name	date_arrival	Month	Year	Arrival	MODAL	Unnamed: 10
0	Karnataka	Mysore	T. Narasipura	Pachha Bale	Fruits	2019-06-18	6	19	6	2880	NaN
1	Karnataka	Mysore	T. Narasipura	Pachha Bale	Fruits	2019-06-11	6	19	10	29	NaN
2	Karnataka	Mysore	K.R.Nagar	Medium	Fruits	2018-08-29	8	18	2500	13000	NaN
3	Karnataka	Mysore	K.R.Nagar	Medium	Fruits	2018-08-17	8	18	1500	13000	NaN
4	Karnataka	Mysore	K.R.Nagar	Medium	Fruits	2018-07-28	7	18	1500	13300	NaN

Fig 4.3 shows the dataset considered for the prediction.

Feature Selection: The subsequent stage involves identifying the pertinent features that exert the most substantial influence on crop prices. Several techniques can be employed for this purpose, including correlation analysis, principal component analysis (PCA), and mutual information.

Train-Test Split: Following feature selection, the dataset is partitioned into training and testing subsets. The training dataset is utilized to train the Random Forest model, whereas the testing dataset serves to assess the model's performance.

```python
e_distName = LabelEncoder()
features.district_name = e_distName.fit_transform(features.district_name)
e_marketName = LabelEncoder()
features.market_center_name = e_marketName.fit_transform(features.market_center_name)
e_variety = LabelEncoder()
features.Variety = e_variety.fit_transform(features.Variety)
e_grpName = LabelEncoder()
features.group_name = e_grpName.fit_transform(features.group_name)
```

```python
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(features,targets, test_size = 0.20)
```


IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁘ Impact Factor 8.102 ⁘ Peer-reviewed / Refereed journal ⁘ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125197

Fig 4.4 shows the Train-Test Split of the dataset.

Model Training: The Random Forest model is trained using the training dataset. This process entails building multiple decision trees by utilizing random subsets of the data. The individual trees are then combined to generate a final prediction.

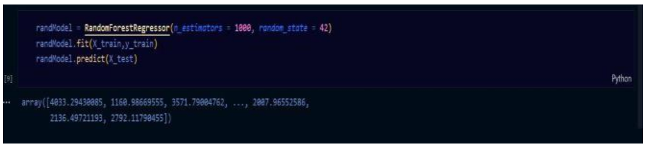


Figure 4.5 provides a depiction of the Random Forest Regressor.

Model Evaluation: The performance of the trained model is assessed using the testing dataset. Various evaluation metrics, including mean squared error, mean absolute error, or R-squared, can be employed to gauge its effectiveness.

Hyperparameter Tuning: To optimize the model's performance, the hyperparameters of the algorithm need to be tuned. This involves adjusting the number of decision trees, the number of features used in each decision tree, and the depth of the decision trees.

Prediction: After the model has been trained and evaluated, it becomes capable of predicting future crop prices. This process entails inputting new data, such as current weather conditions and economic indicators, and utilizing the trained model to generate a prediction.

Model Interpretation: After training and evaluating the model, it becomes capable of predicting future crop prices by inputting new data, including current weather conditions and economic indicators, and utilizing the trained model to generate predictions.

5. CONCLUSION:

The Raitha Saathi Application proposes a system that harnesses the powerful capabilities of the Random Forest algorithm for crop price prediction. This algorithm constructs multiple decision trees by utilizing random subsets of the data and features, and then combines their individual predictions to generate a final prediction. Random Forest is particularly suitable for crop price prediction as it effectively mitigates overfitting and can handle both categorical and numerical data.

The process of building an accurate and reliable model for predicting future crop prices involves several steps. These include data collection and pre-processing, selection of relevant features, model training and tuning, and evaluation of its performance. By following these steps, a robust model can be developed to provide reliable predictions for future crop prices.

The utilization of Random Forest in crop price prediction has significant benefits for farmers, traders, and policymakers. It empowers them to make well-informed decisions and enables effective future planning. Ultimately, this contributes to a more efficient and sustainable agricultural industry.

REFERENCES

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁝⁝ Impact Factor 8.102 ⁝⁝ Peer-reviewed / Refereed journal ⁝⁝ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125197

[1]Yung-Hsing Peng, Chin-Shun Hs, and Po-Chuang Huang, "Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets," IEEE 2017.

[2] Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty.2017. Agricultural Production Output Prediction Using Supervised Machine Learning Techniques.978- 1-5386-3831-6/17/$31.00 ©2017 IEEE.

[3] S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh.2019. Machine learning approach for forecasting crop yield based on climatic parameters .978-1-4799-2352- 6/14/$31.00 ©2014 IEEE

[4] Neha Rale, Raxitkumar Solanki, Doina Bein, James Andro-Vasko, Wolfgang Bein. Prediction of Crop Cultivation. 978-1-7281-0554-3/19/$31.00©2019 IEEE.

[5] G Srivatsa Sharma, Shah Nawaz Mandal, Shruti Kulkarni, Monica R Mundada, Meeradevi.2018. Predictive Analysis to Improve Crop Yield Using a Neural Network Model. 978-1- 5386-5314- 2/18/$31.00 ©2018 IEEE.

