Journal of Information Technology and Digital World (ISSN: 2582-418X)
www.irojournals.com/itdw/

Krishi-Stats: A Web-based System for Crop Price Prediction using Machine Learning Approach

Dakhole Dipali
Computer Science and Engineering, Presidency University, Bangalore, India
E-mail: prof.dipali@gmail.com

Abstract

Agriculture is the main livelihood in India. Most of the people earn bread and butter through farming, but the farmers are not getting enough profit and the field is facing growth downward due to irregular rainfall, high volatility in agriculture commodity prices and uncertainties in production. The objective of this study is to design and implement an automated crop price prediction system with best suitable machine learning technique, as well as displaying prediction results on website Krishi-Stats designed for easy understanding for Farmers. In this study, three machine-learning (ML) algorithms, ARIMA, VAR and XGBoost are applied on large historical data collected from government website. The ML algorithms compared with their root mean square error values (RMSE). As XGBoost has given optimum RMSE value of 0.94, has been selected as the prediction system engine of our website Krishi-Stats. On website, the crop prediction prices are plotted for all twelve selected crops and visualized using prediction graphs.

Keywords: Precision agriculture, crop price prediction, machine learning, XGBoost, ARIMA, VAR.

1. Introduction

In India, majority of people has source of income as farming. The agriculture sector contributes to 17.32% of the GDP of the country [7]. Most of the land in the country is used for agriculture to meet the needs of growing population. However, the farmers in the country are struggling with profit and prices of the product produced in their farms due to uncertainties in the production, high volatility in the prices of agricultural commodities and frequently changing climatic conditions. Most of the times, farmers are not getting profit due

DOI: https://doi.org/10.36548/jitdw.2022.3.006
Received: 07.08.2022, received in revised form: 21.08.2022, accepted: 06.09.2022, published: 16.09.2022
© 2022 Inventive Research Organization. This is an open access article under the Creative Commons Attribution-NonCommercial International (CC BY-NC 4.0) License

to low prices, in spite of good production. Similarly, consumers have to pay excessive prices. One of the major reason for this is the lack of technology in agriculture field.

Keeping all this in mind, a tech platform Krishi-Stats for crop price prediction with help of machine ML is created to make technology use in agriculture, which will benefits to the farmers. Traditionally the price prediction in the country used to happen by considering some set/sample population and corresponding crop production, and used to be applied for whole population in all regions of the country [12]. Price forecast is one of the critical inputs to the farmers to take the production and marketing decisions as well as to the agricultural administering authorities to make the policies [3].

Machine learning has given successful prediction models in various fields such as stock market, weather prediction, business decisions, online marketing, and yield prediction as well as in crop price predictions [15]. There are two ways to apply machine-learning models, supervised and unsupervised learning. If the data is labeled, supervised learning is applied to do the prediction; if unlabeled data is available then unsupervised algorithms are used to find hidden patterns of prediction [8]. To predict the price of Arecanut, SARIMA, Holt Winter's seasonal method and LSTM neural network were used and LSTM has given better performance in terms of root mean square error (RMSE) value [7]. The study forecast interval-valued agricultural commodity predicted prices with vector error correction model (VECM) and multi-output support vector regression (MSVR), VECM is capable of capturing both the linear and nonlinear patterns implemented to calculate agricultural commodity forecast prices. The proposed model has given expected results to forecast the future commodity prices for cotton and corn [16].

The five different prediction algorithms were applied for cotton price prediction and their RMSE, Mean Square Error (MSE), Relative Absolute Error (RAE) and Relative Standard Error (RSE) values were calculated. The Boosted Decision Tree algorithm is given better performance than Linear Regression (LR), Bayesian Linear Regression (BLR), Decision Forest Regression (DFR) and Decision Forest Regression (DFR) [13]. Auto Regressive Integrated Moving Average (ARIMA) Model was programmed using R language to predict cotton future prices; the performance of the model is examined by Akaike Information Criterion (AIC), and Minimum Absolute Percentage Error (MAPE) [5]. A web based automated system is implemented to predict agricultural commodity prices, for prediction five ML strategies have been compared. LSTM has given better results than ARIMA, XGBoost, Prophet, and SVM; hence, it is selected as prediction engine on web [17].

ARIMA technique is used to predict the future prices of onion. To finalize accurate values of number of seasonal auto regressive terms (P), number of seasonal differences (D) and number of seasonal moving average terms (Q), different ARIMA models have been designed and compared their R-square, Mean Absolute Percentage Error (MAPE), RMSE and MAE [9]. Different ARIMA time series models were implemented and estimated to forecast wheat prices. ARIMA with P=1, D=1 and Q=1 model was best fitted and correctly predicted the future wheat prices [14].

The prices of different crops is predicted using rainfall data and Wholesale Price Index (WPI) data using supervised decision tree regression machine ML model [6]. The paper has employed two approaches thin-tailed normal distribution and the fat-tailed student t-distribution model for wheat, rice, beef, groundnut, sugar and coffee. The t-distribution model has given better performance than normal distribution [10]. In [1], Autoregressive Integrated moving average (ARIMA) and Generalized Auto Regressive Conditional Heteroscedastic (GARCH) were used for crop price prediction, in which GRACH has given better performance than ARIMA.

The study shows that, ML has been applied in agriculture to discover the knowledge as well as insights into it. Various ML approaches are used to predict crop prices on different data sets. Every time a different ML technique is giving better performance than other ones [11]. Hence, we can't say that a particular ML technique best suited for crop price prediction. The accuracy of model depends on the data set selected as well. In given study ARIMA, VAR, and XGBoost models are used for crop price prediction, as they are time series models. In addition, this work aimed to design and implement a website Krishi-Stats using best fit machine learning algorithm for crop price prediction. Total twelve crops Bengal gram dal, Avaredal, Tur dal, Chana dal, Wheat, Ragi, Rice, Groundnut, Onion, Maize, Sugar, and Potato datasets were selected for prediction results to display in the form of graphs using python libraries on website. The proposed system can support farmers to find future prices of these twelve crop for any period and helps them to plan for correct selection of crops for correct period to gain profit more.

2. Materials and Methods

In this work, a website called Krishi-Stat, is designed and implemented for crop price prediction using ML techniques. ARIMA, VAR, and XGBoost are the time series ML techniques, used and compared for predicting different crop future prices. These models are capable of forecasting crop future prices with trends and seasonality.

2.1 ARIMA

ARIMA (Auto-Regressive Integrated Moving Average) is a time series forecasting model and variation of Box Jenkins models. ARIMA technique has three parameters to set, P number of seasonal auto regressive terms, D number of seasonal differences and Q number of seasonal moving average terms, denoted as ARIMA (P, D, Q) [4]. Based on P, D and Q values, the model chosen to check the performance of ARIMA. ARIMA(1,1,1) has given best result and selected for prediction.

2.2 VAR

A vector auto regression (VAR) involves k random variables, which gives k different equations. It is simple modeling method. All variables are root and the Ordinary Least Squares technique (OLS) can be applied for estimation of parameters. These parameters make it advantageous over other multivariate modeling techniques like simultaneous equation models [2].

2.3 XGBoost

XGBoost is known as Extreme Gradient, which implements bagging and boosting methodology. This model pushes the limit of computing power for boosted trees algorithms and is scalable to situations because of its algorithmic optimizations. It reduces time complexity as well as space complexity. It is ten times faster than ARIMA, SVR, Prophet and Long Short-Term Memory (LSTM) on a single machine [17].

2.4 System Overview

The web system works as shown in Fig.1. The design follows Model-View-Controller-Structure (MVC). On back end of website, the prediction engine works with trained data and applies XGBoost ML techniques for crop price prediction. The application server is the price prediction engine, which applies XGBoost technique on the connected dataset. The server processes the request send, and the graphical view of crop price prediction is displayed on forecast page of website. The crop price prediction graph has plot for the original price and predicted price for comparison. The website is designed using HTML and JavaScript coding. For integrating XGBoost into website python libraries NumPy, Pandas, Sklearn, Flask, PymySql are used.

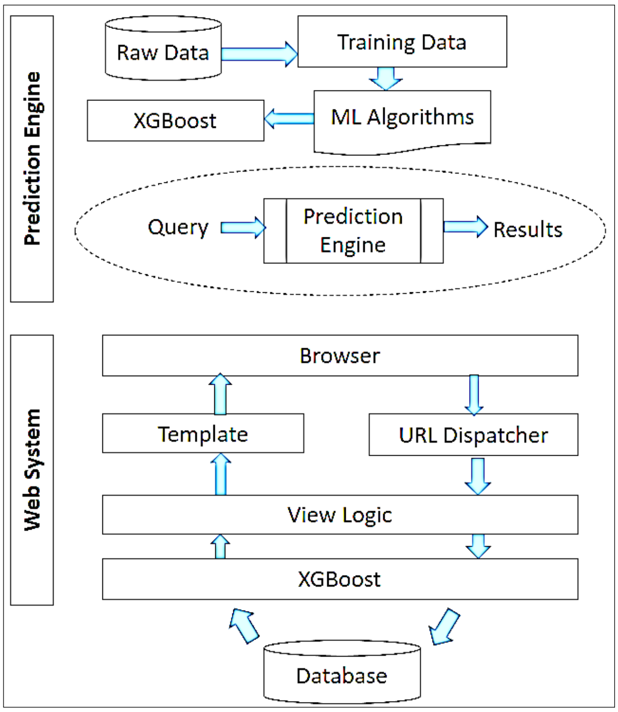


Figure 1. System Overview

2.5 Prediction Engine Design

The purpose of prediction engine is to implement XGBoost in python code and get connected with dataset while extracting prediction results. In prediction engine, there are four steps. In first step, the data is collected from database from government website. In second step, the collected data is cleaned using Python libraries NumPy and Pandas to maintain homogeneity. In third step the model building is done in Python. In fourth step, the best-fit model XGBoost is selected to visualize the results on web Krishi-Stat.

3. Experimental Setup

The main focus of the project is forecasting. Hence an experiment has been done to find the optimal algorithm which performs forecasting with high accuracy. From the study, three algorithms ARIMA, VAR, and XGBoost were selected as the potential prediction system engine by comparing their performance to learn the best model to fit from large datasets.

3.1 Data Set

The data set is extracted from government website https://agmarknet.gov.in/. The website contains data for various agriculture commodities. From these commodities Bengal gram dal, Avaredal, Tur dal, Chana dal, Wheat, Ragi, Rice, Groundnut, Onion, Maize, Sugar, and Potato were selected. Each crop dataset is separate and consists of date, minimum prize, maximum prize, and modal price. The model price is the average of the prices minimum price and maximum price. The dataset is collected from date 25th May 2013 to 25th May 2022. The sample data set in Table 1 is cleaned using Python library functions to maintain homogeneity in dataset.

Table 1. Sample Data Set

Sr. No	Min_Price (Rs./Quintal)	Max Price (Rs./Quintal)	Modal_Price (Rs./Quintal)	Price_Date
1790	4000	4300	4150	5/25/2013
1789	4000	4300	4150	5/27/2013
1788	4000	5100	4600	5/28/2013
1787	3060	4860	4500	5/29/2013
1786	3060	4860	4500	5/30/2013
1785	3800	4800	4400	5/31/2013
1784	3900	4520	4300	6/1/2013

3.2 Configuration of ML Models

All ML models ARIMA, VAR and XGBoost were analyzed and visualized using python IDE. For modelling, the following configuration steps were followed for each ML model.

3.3 Configuration of ARIMA

ARIMA consists of three parameters (P, D, Q) [15]. First to determine the time series is stationary by conducting an Augmented Dickey Fuller Test (ADCF) on the dataset and a graph is plotted. In this plot first negative plot value is value for D, which resulted into ARIMA (D, 1). Next, demonstrate the correlation between the previous time period with the current for all data points. In order to predict the value of P, using Eq. (1), directly plot a partial auto-correlation function (PACF) graph, which resulted into ARIMA (P, 1). To calculate Q, the main purpose is to find out the noises that could potentially affect the model. Hence, plot an auto correlation function (ACF) plot using Eq. (1), which resulted into ARIMA (Q, 1).

$$Y(t) = \alpha + \beta(1).Y(t-1) + \beta(2).Y(t-2) + \cdots + \beta(p).Y(t-p).\epsilon(1) + \varphi(1).\epsilon(t-1) + \varphi(2).\epsilon(t-2) + \cdots + \varphi(q).\epsilon(t-q) \quad (1)$$

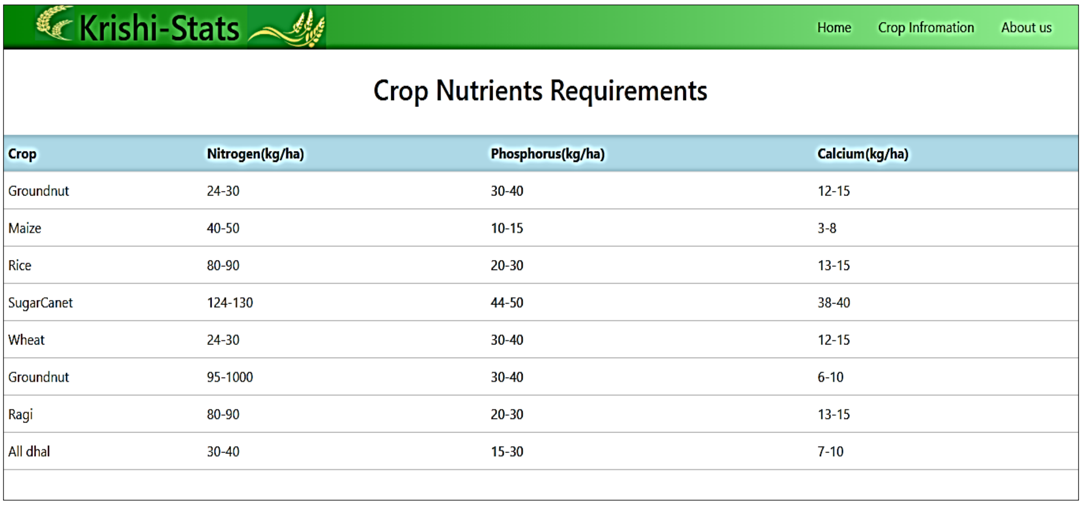
3.4 Configuration of VAR

After loading the dataset, find random variable p using augmented design fuller test to see weather data is stationary or not. If p value is than 0.05, then it is stationary. Check the correlation of the data using Granger Causality Test, and then dividing train, test data set as 8:1, and fit the model to print the summary.

3.5 Configuration of XGBoost

The data was pre-processed first on formatting as well as computed helper variables as well, for the required format so that it can be easily trained and tested in our XGBoost algorithms. After pre-processing of the data, data is again stored in another .csv file. Then the saved CSV data was loaded into the XGBoost algorithm. After loading the dataset, it has been divided into testing and training using the train_test_split with split ratio of 8:1. To optimize the model, XGB Regressor class with hyperparameters is used. It fits the training data and generate predictions on the testing data. Finally, using Plotly library the original and forecasted test results are shown in plots. The code is uploaded into the flask server. The visual studio code is run in the terminal where the link for opening the website is received.

4. Website Implementation



Crop	Nitrogen(kg/ha)	Phosphorus(kg/ha)	Calcium(kg/ha)
Groundnut	24-30	30-40	12-15
Maize	40-50	10-15	3-8
Rice	80-90	20-30	13-15
SugarCanet	124-130	44-50	38-40
Wheat	24-30	30-40	12-15
Groundnut	95-1000	30-40	6-10
Ragi	80-90	20-30	13-15
All dhal	30-40	15-30	7-10

Figure 2. Crop Information

Fig. 2. Krishi-Stats Crop Information Our website Krishi-Stats is designed for providing a statistical time series analysis of the crop price. The purpose of designing this website is to benefit the farmers to cultivate which crop at which time duration, to gain more profit in future. The website helps farmers to get future prices of crop as well crop nutrient

information. Here we have three web pages that are designed for website. The website is made simple but useful for farmers.

4.1 Home Page

On home page, twelve crops Potato, Onion, Bengal gram dal, Avare dal, Rice, Wheat, Ragi, Chana dal, Sugar cane, Groundnut, Maize, and Tur dal have been enlisted as shown in Fig. 3. On clicking in their image icons, the complete crop price prediction graph until 2075.

Figure 3. Krishi-Stats Home Page

4.2 Crop Information

On the Crop information page, the user/farmer can see what all crops need and how much amount of nitrogen, phosphorus, and calcium is needed for the particular crops as shown in Fig. 3.

4.3 About Us

The about us page displays that in what aim the Krishi-Stats website has been developed and how it will help farmers and what is the role of the Krishi-Stats website for the farmer economy. This page gives information and vision of website designed.

5. Results and Discussion

5.1 Comparison between ARIMA, VAR and XGBoost

All three ML models have been evaluated on comparing their Root Mean Square Error (RMSE) value as shown in Table 2. The RMSE value obtained for all crop using ARIMA model is very high, hence it shows that for this application ARIMA is not suitable as it is suitable for single variant applications. The RMSE value results for VAR is still greater than 1 and comparatively high than that of XGBoost. Hence, XGBoost is best suited for crop price prediction over the dataset used. The RMSE value for all crops is in between 0.8 to 0.99 using XGBoost. Finally, as XGBoost have given better results for price prediction of each crop, it has been selected and integrated in prediction engine of website.

Table 2. RMSE values for all crops for ARIMA, VAR and XGBoost Models

Crops	ARIMA	VAR	XGBoost
Potato	3472.75	1.4	0.8
Onion	3861.86	1.3	0.96
Bengal gram dal	12560.73	1.4	0.97
Avarie dal	14079.80	1.5	0.96
Rice	140381.52	1.6	0.87
Wheat	63573	2.3	0.87
Ragi	153686	1.4	0.94
Channa dal	31715.78	1.7	0.98
Sugar cane	13523	0.98	0.96
Groundnut	46173	1.6	0.98
Maize	39886	1.9	0.97
Tur Dal	205276.8	2.5	0.99

5.2 Price Prediction Results for each Crop

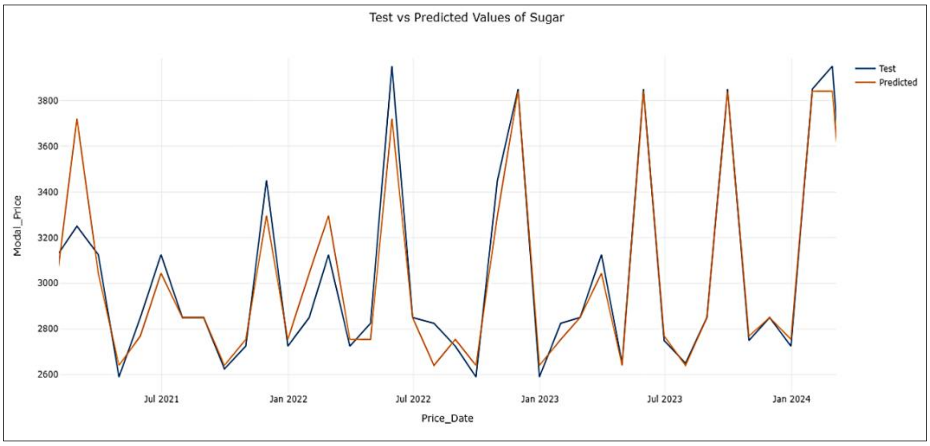


Figure 4. Sugar Price Prediction Graph

Extreme Gradient (XGBoost) is used for crop price prediction selected for website designing, as it has given better performance than ARIMA and VAR model. Fig. 4 is the graphs that illustrate the original prices, as well as the predicted prices using XGBoost algorithm for 12 experimental commodities (Potato, Onion, Bengal gram dal, Avare dal, Rice, Wheat, Ragi, Chana dal, Sugar cane, Groundnut, Maize, Tur dal). The graph helps us to visualize the outcome of predicted price in order to make a visual presentation to users. Each graph has date on x-axis and model price on y-axis. The original price plot and predicted price plot are plotted on same graph for comparison. Framers/Users can select any point on this graph to check for future price of crop at particular day/date/duration. This will give an idea for framers to cultivate which crop at which time slot to gain more profit in future.

6. Conclusion

The study is aimed to aid agriculture by designing and implementing a website Krishi-Stats on the crop price prediction. The study shows that there are various techniques are used for crop price prediction including some ML approaches. ARIMA, VAR and XGBoost are the ML techniques used in this work for crop price prediction. On comparing their RMSE value, XGBoost has given better performance with RMSE value varying between 0.8 to 0.99. XGBoost is a multi-variate time series and is a high-level forecasting ML algorithm for any statistical projects. Hence, it is used for prediction engine in Krishi-Stats website.

In Krishi-Stats website, twelve crops Potato, Onion, Bengal Gram Dal, Avarie Dal, Rice, Wheat, Ragi, Channa Dal, Sugar Cane, Groundnut, and Maize are chosen on the home page for forecasting and the result are displayed graphically on a new page. In each graph the prediction results are shown up to 53 years. In future more features like crop recommendations based on soil types and contents can be added to make it more useful for the users.

7. Acknowledgments

The authors of this publication declare there is no conflict of interest. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] Bhardwaj, S. P., Paul, R. K., Singh, D. R., & Singh, K. N. (2014). An empirical investigation of ARIMA and GARCH models in agricultural price forecasting. Economic Affairs, 59(3), 415.

[2] B S YASHAVANTH, K N SINGH, AMRIT KUMAR PAUL, RANJIT KUMAR PAUL, Forecasting prices of coffee seeds using Vector Autoregressive Time Series Model, Indian Journal of Agricultural Sciences, January 2017.

[3] Brandt, J. A., & Bessler, D. A. (1983). Price forecasting and evaluation: An application in agriculture. Journal of Forecasting, 2(3), 237-248.

[4] Chandran, K. P., & Pandey, N. K. (2013). Potato price forecasting using seasonal ARIMA approach. AGRIS, 0970-8235.

[5] Darekar, A., & Reddy, A. A. (2017). Cotton price forecasting in major producing states. Economic Affairs, 62(3), 373-378.

[6] Dhanapal, R., AjanRaj, A., Balavinayagapragathish, S., & Balaji, J. (2021, May). Crop price prediction using supervised machine learning algorithms. In Journal of Physics: Conference Series (Vol. 1916, No. 1, p. 012042). IOP Publishing.

[7] Kiran M. Sabu, T.K. Manoj Kumar, Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala, Procedia Computer Science, Volume 171, 2020, Pages 699-708, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.04.076.

[8] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674.

[9] Mulla, Areef. (2020). Price Behavior and Forecasting of Onion Prices in Kurnool Market, Andhra Pradesh State. Economic Affairs. 65. 10.30954/0424-2513.1.2020.6.

[10] Onour, I., & Sergi, B. S. (2011). Modeling and Forecasting Volatility in the Global Food Commodity Prices (Modelování a Prognózování Volatility Globálních cen Potravinářských Komodit). Agricultural Economics-Czech, 57(3), 132-139.

[11] Pham, X., & Stack, M. (2018). How data analytics is transforming agriculture. Business horizons, 61(1), 125-133.

[12] Rakhra, M., Soniya, P., Tanwar, D., Singh, P., Bordoloi, D., Agarwal, P., & Verma, N. (2021). Crop Price Prediction Using Random Forest and Decision Tree Regression:-A Review. Materials Today: Proceedings.

[13] S. Deepa, A. Alli, Sheetac, S. Gokila, Machine learning regression model for material synthesis prices prediction in agriculture, Materials Today: Proceedings, 2021, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.04.327.

[14] Sharma, H. (2015). Applicability of ARIMA models in wholesale wheat market of Rajasthan: An investigation. Economic Affairs, 60(4), 687.

[15] SUSSY BAYONA-ORÉ, RINO CERNA2 , EDUARDO TIRADO HINOJOZA2, “Machine Learning for Price Prediction for Agricultural Products” WSEAS

TRANSACTIONS on BUSINESS and ECONOMICS, ISSN: 2224-2899, Volume 18, 2021.

[16] Xiong, T., Li, C., Bao, Y., Hu, Z., & Zhang, L. (2015). A combination method for interval forecasting of agricultural commodity futures prices. Knowledge-Based Systems, 77, 92-102.

[17] Zhiyuan Chena, Howe Seng Goh, Kai Ling Sin, Kelly Lim, Nicole Ka Hei Chung, Xin Yu Liew, Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques. Advances in Science, Technology and Engineering Systems Journal, Volume 6, Issue 4, Page No 376-384, 2021.

Author's biography

Dipali Dakhole is currently working as Assistant Professor at Presidency University, Bengaluru, India. She obtained a graduate degree in CE from University of Pune, India. She received a postgraduate degree in WCC from the Department of CSE, G. H. Raisoni College of Engineering, Nagpur, India. She worked as an Assistant Professor in PRMIT&R Engineering College, Amaravati India in 2015. She also worked in ISB&M School of Technology, Pune as an Assistant Professor from May 2016 to May 2017. Her research interest includes control systems, decision theory and internet of things.