International Journal of Agriculture Innovations and Research
Volume 8, Issue 3, ISSN (Online) 2319-1473
Manuscript Processing Details (dd/mm/yyyy):
Received: 22/10/2019 | Accepted on: 09/11/2019 | Published: 21/11/2019

Centralized System to Bridge the Gap between Supply and Demand of Agro-Products using A.I. and Data Analysis

Dhruva Phulsundar [1*], Mitalee Minde [2], Saurabh Shingte [3], Shubham Gujare [4] and Dr. S. L. Vaikole [5]
[1, 2, 3, 4] Department of Computer Engineering, Datta Meghe COE, Airoli, Navi Mumbai, India.
[5] Associate Professor, Datta Meghe COE, Airoli, Navi Mumbai, India.
*Corresponding author email id: dhruv.phulsundar@gmail.com

***Abstract* – India is an agrarian country with more than two thirds of its population depending directly or indirectly upon agriculture, thousands of crores of rupees are spent every year on farmers through various channels by the government because of inadequate remuneration received by the farmers. One of the major reasons being the lack of technological solutions in the field of agriculture, especially in a developing country like India. Bridging the gap between the demand and supply using technology for all the agro based products will help the farmers as well as the government to make effective decisions. Based on the complexity of prediction by making use of the Random forests machine learning method to its relatively good accuracy, robustness and ease of use for building up a model to eliminate Minimum Support Price a Booking mechanism is developed which helps to keep track of the demand i.e. the need of the market for a particular agro product, and supply i.e. the amount of products that shall be delivered to the market that can be controlled via a mobile application by the farmers for the farmers. The image processing model provides security and helps with the validation of the users. The Mapping function helps the government and farmers in tracking, locating and getting real-time updates of the farm lands being cultivated.**

***Keywords* – Minimum Support Price, Random forests, Image processing, Machine learning, APMC.**

I. INTRODUCTION

Indian Agriculture being the backbone of the economy contributes to the overall economic growth of the country and determines the standard of life for more than 60 percent of our people [1]. India pioneers in the food and grocery market is making it the sixth largest in the world, with retail contributing 70 % of the sales [2]. Agricultural productivity depends on several factors such as availability and quality of agricultural inputs such as land, water, seeds and fertilizers, access to agricultural credit, crop insurance along with assurance of remunerative prices for agricultural produce, and storage and marketing infrastructure, among others [3].

A. *Motivation*

Due to the lack of technological solutions in this field there is no system in place to keep the farmers updated in real time. Insufficient knowledge about the demand and supply of agro products, farmers usually decide to produce a particular set of crops just by listening to other farmers or by falling for the word of mouth publicity resulting in over production of a single crop and thus not getting any profit out of it when delivered to the market. As the demand of these agro based products being delivered to market changes moderately, the supply can be streamlined as per the demand (requirement of the market). The main aim of the system is to analyse the need of the market and amount of products being supplied to the market, calculate the difference and bridge the demand supply gap by using Machine Learning techniques. The system will perform calculations and display these details in a simple graphical format to the farmer. The data will dynamically change as per the user's (Farmer) inputs such as the selection of the crop to be cultivated, land details, location, approximate amount of yield, nearest

APMC (Agricultural Produce Market Committee) market and projected MSP. The system will mainly target following topics.

Bridging the Gap between Demand and Supply:

The system will know the rate of consumption i.e. the demand of a particular crop at a particular time thus dynamically updating the figures on the website in a graphical format. The farmers can then easily identify and analyse the requirement of a particular crop and plan their set of actions by booking their delivery in advance. The system will perform mathematical calculations and keep updating the overalls as and when a slot is booked by a farmer. Overall, bridging the gap of supply and demand.

Wastage of Food Due to Overproduce:

The excessive production of any agro product will come under control, once the supply is met with the demand of a particular product for a particular month. The system will notify the farmer about the same by suggesting an alternative. Thus eliminating the food wastage due to excessive production.

Lack of Minimum Support Price on crops:

The supply and demand will be matched by the system. Once the cycle fits in, a lower bound price can be projected by the help of machine learning technique which will act as an assurance to the farmers.

Farmers' suicides in India:

The system aims to eliminate farmer suicides which is a major issue currently faced by the country. Since the MSP can be projected for each product, farmers will act smartly and won't suffer a loss on their production. Which will help them repay their loans. A win-win situation for both farmers and the government.

Proper Channelising of Crop Insurance by the Government:

The system will have records of all the farmer's registrations and bookings details of agro products. Thus in case of a natural calamity it would be easy for the government to track and narrow down the affected areas under cultivation and reimburse the eligible farmers their money.

Efficient Tracking and Live Updates of the Lands:

The system will provide geographical data with the help of maps, details of active plantation of a particular crop as well as mapped locations of the APMC markets, making it an efficient tool for the government to manage, analyse and track the activity status of crops in real time.

II. METHODOLOGY

A. *Image Processing*

User information authentication is a major factor in developing a user-centric web-portal. By using image processing we reduce the number of clicks user has to undergo while registration. Doing so, one avoids unnecessary mistakes while typing in the data. Also this helps differently-abled people to fill in the data.

In this module, Tesseract [4] OCR (Optical Character Recognition) is being used. Tesseract is one of the OCR engines available. It has functionality which can be implemented for unicode and reading more than 100 languages

[5]. Training can be done to read other languages also.

User's Aadhaar Card is sent to ORC and user details like name, address and date of birth are fetched from it.

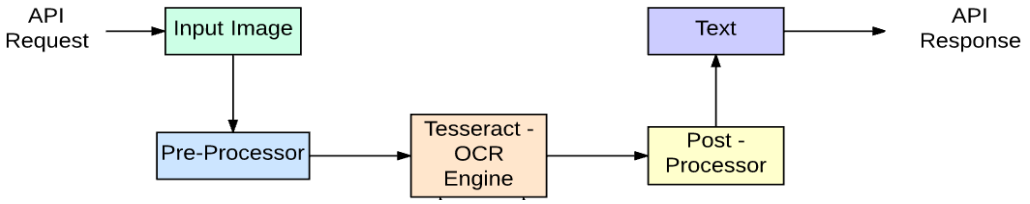


Fig. 1. OCR Process Flow [12].

B. *Slot Booking*

Failure in crop production is not the actual trigger of Farmer's crises. Market volatility is the current bane. Before growing any crops farmers consider the season, vegetation and the geographical location, but none of them take market conditions into consideration. Since there is no technological solution to provide Farmers with a platform to view what is going to be supplied to the market and in what quantity. Therefore at times there may be overproduction or underproduction of a certain crop.

As supply of the crops plays an important role in deciding MSP. For this Booking mechanism has been developed. This mechanism asks the farmer to book a delivery slot of the upcoming produce by entering details such as:

- Selecting the crop.
- Selecting the date of harvest for the crop.
- Entering the approximate delivery quantity of the produce.

Once the user enters the above details, Booking Mechanism will book the slot for that user. Progress of each booking will be shown in my bookings page in the below manner.

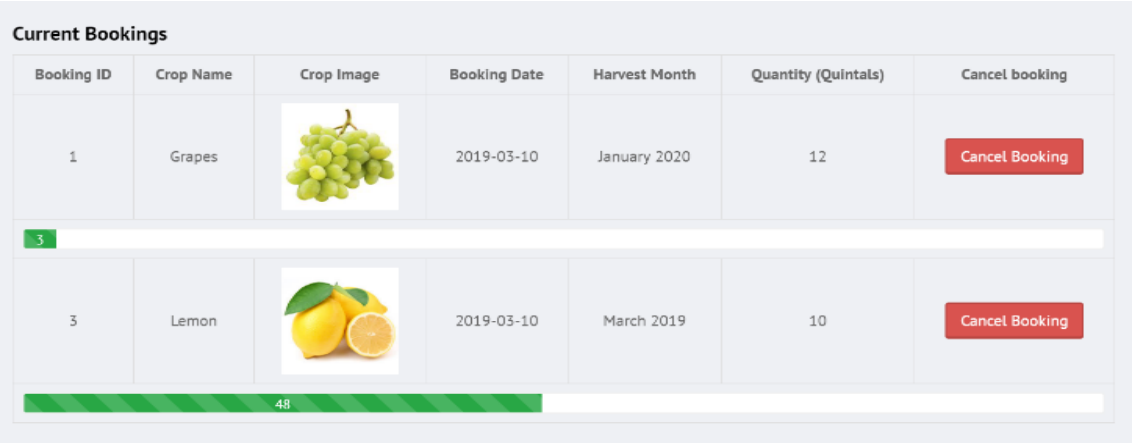


Fig. 2. Booking Mechanism [13].

Where user can view the current bookings, progress and manage or modify bookings. User can also view previous bookings.

C. *Visualization using Maps*

The purpose of this project is to help farmers and government officials to analyze and track the agriculture produce with the help of statistical data. This module is implemented with the help of Google maps. Google maps is one of the biggest and open source map services which provides views and can handle large amount of data [6]. This feature allows the farmers to view the following.

- Nearby APMC markets and/or around a specified region.
- Geographically locating any particular crop that is being cultivated.

Farmers are very less informed about the closest. APMC markets around them, which results in a higher amount of transportation cost. This problem can be solved by this module. Farmers just have to select a location from the options and the maps point out the APMC'S around that region.

In order to cultivate a specific crop or vegetable or fruit one needs to analyze the ongoing crop production of that very crop so as to avoid excess production. In order to tackle this issue, farmers can use the map feature. When a crop is selected from the options, the maps dynamically updates and pinpoints all the locations where the selected crop is currently being cultivated.

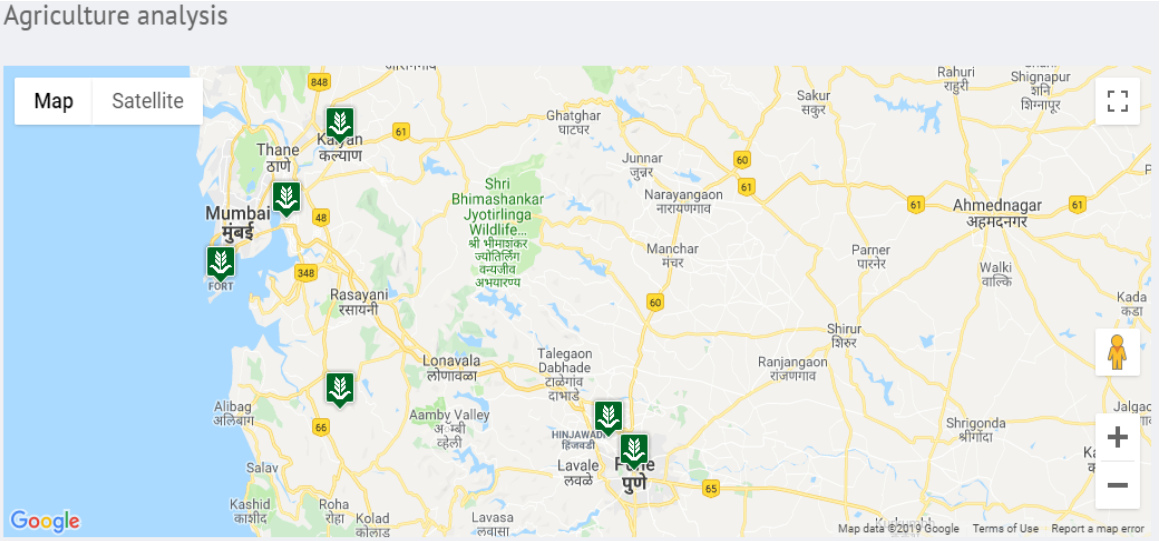


Fig. 3. Map Visualization [13]

D. *Multi-Language Translation*

A large number of people in India do not know English. To make the system more user friendly, there is a feature of language translation. Using Google Language Translation the system can translate to and fro in languages supported by Google [7].

E. *Prediction Mechanism*

Random forest algorithm is one of the supervised classification algorithms. Random Forest Algorithm proceeds with creating number of trees. In general, the robustness of forest depends upon the number of trees generated. That is the accuracy of random forest classifier increases as the number of trees increases. As all the calculation of nodes selection will be same for the same dataset.

Decision tree concept is built on a system based on rules. After training on the dataset, the decision tree algorithm will provide some rules. These rules will be used for prediction or classification. Suppose a model is to

be trained to predict whether a person would like your food or not. The model will look for the characteristics of the food he ate in the past, such as temperature of food, spiciness, saltiness etc.

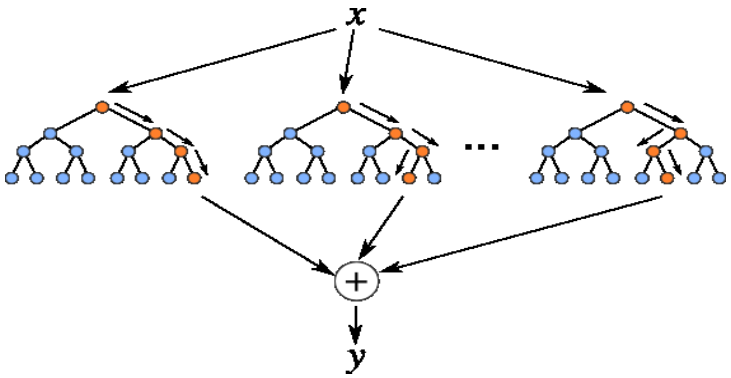


Fig. 4. Random Forest Tree [14].

F. *Advantages of Random Forest*

- This algorithm can be used for classification as well as regression purpose.
- Missing data is well handled.
- Model does not overfit even if large number of trees are used.
- Time taken is very less compared to other algorithms [8].

G. *Implementation*

Random Forest is one of the supervised learning algorithms. It creates a forest of trees which are generated randomly [9]. The forest is nothing but an ensemble of trees having some predicted value, which is trained using bagging method [9].

1. *Bagging*

The general method of bagging is used for the training purpose of the random forest regressor [10]. Consider a training set $X = x_1, ..., x_n$ having outputs $Y = y_1, ..., y_n$, repeatedly bagged for times, fits trees to randomly selected samples (by bagging) [10]:

For $b = 1, ..., B$:

1. N samples from dataset X, Y as X_b, Y_b with replacement[10].
2. On X_b, Y_b a regression tree is trained [10].

From all the individual trees on x', prediction for any unseen sample x' can be made [10]:

$$f = \frac{1}{B} = \sum_{b=1}^{B} f_b\,(x') \tag{1}$$

Bootstrapping is one of the methods to decrease the variance of the model. This method is more useful because it doesn't increase the bias [10]. That is even if the prediction of each tree is highly variable to the dataset, the average of many trees is not [11]. Bootstrap sampling decorrelates trees which are highly correlated trees [10]. Condition of highly correlated trees happens if we use only one dataset for training.

Moreover, standard deviation of the predictions of every tree on x' can be considered as an estimate of the prediction [10].

$$\sigma = \sqrt{\frac{\sum_{b=1}^{B} (fb(x') - \overline{f})^2}{B - 1}} \quad (2)$$

2. *Proposed Model*

Proposed Model requires prediction of future data. Since, there are two values to be predicted in the model, the random forest regressor is used two times in it. The two values are demand and MSP (rate) for some months in the future. Two different regressors are used for the prediction purpose.

Before executing the code, admin has to add the time till where the prediction is required. Accordingly, zeros or null are added in the adjacent columns. On execution algorithm first train both the regressors with appropriate features. Then it looks for the zero or null value. If it is in demand column it calls 1st regressor and starts predicting demand data. On completion, it invokes 2nd regressor and starts predicting msp.

1st Regressor is trained with the timeline (months) to predict the monthly demand. That is, the model is trained with only one feature. So, the model predicts demand for next months based on previous months. Prediction will be more accurate if more features are added such as rainfall, crop disease data.

2nd Regressor is trained with two features. Timeline (months) and the demand data are the features used for the prediction of msp (rate). This MSP will serve as an important parameter for farmers to decide whether to plant the crop or not. Because of two parameters less error in the prediction is seen.

III. Experimental Results and Performance Analysis

A. *Comparison between ANN and RFR*

The model is trained using different algorithms of machine learning. Such as ANN (Artificial Neural Network) and RFR (Random Forest Regressor). By comparing different results, RFR seems to be the best fit for the model.

1. *ANN Results*

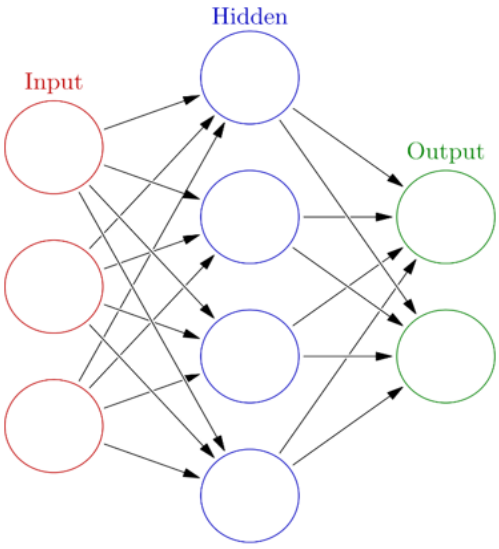


Fig. 5. Artificial Neural Network (ANN) [15].

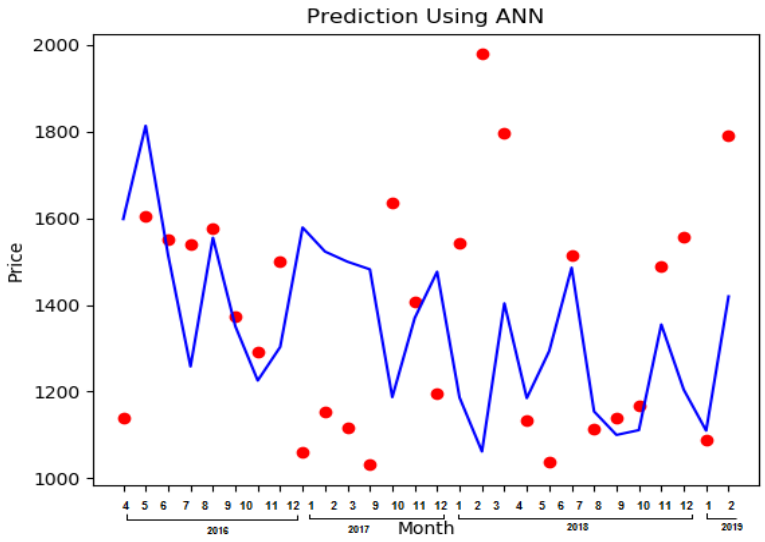


Fig. 6. Monthly MSP Prediction for onion using ANN.

Table 1. ANN Result (Analysis of Fig 6).

Month	Actual Value	Predicted Value	Error
1	1979	1272.486	35.70
2	1797	1420.3392	20.96
3	1133	1198.4161	5.77
4	1039	1209.1298	16.37
5	1516	1505.0778	0.72
6	1114	1168.3671	4.88
7	1138	1113.5393	2.15
8	1168	1125.21	3.66
9	1489	1373.2806	7.77
10	1557	1220.8212	21.59
11	1087.5	1125.0931	3.46
12	1791	1440.3585	19.58

From fig. 6. and table 1 one observes that there are following problems while using ANN model:

1. Prediction contains too much error.
2. Peak values are often neglected.
3. Little bit slower than other algorithms.
4. Requires more memory space.
5. Error Percentage - 59%.

2. *RFR Algorithm Results*

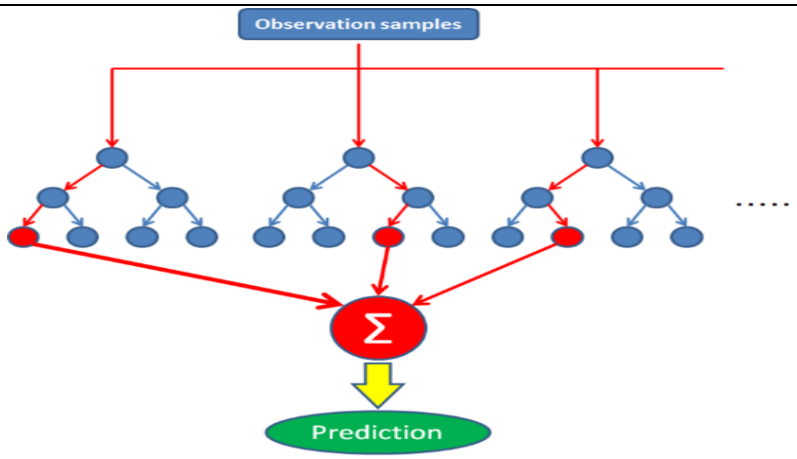


Fig. 7. Random Forest Model.

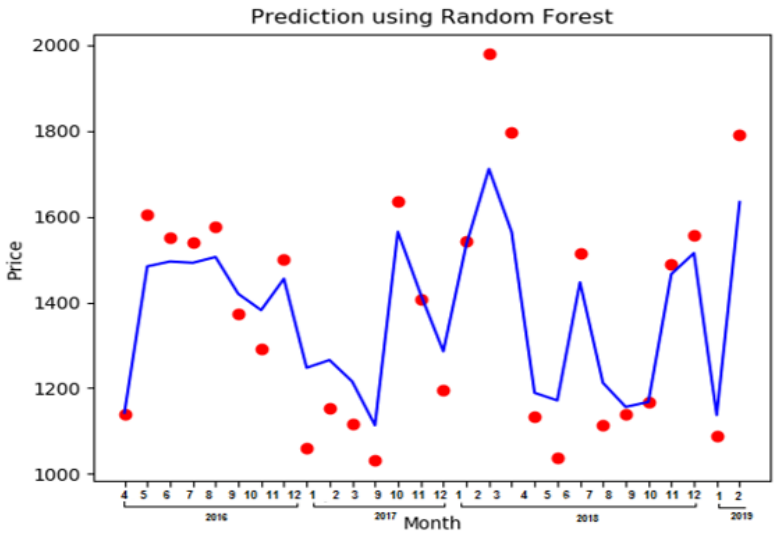


Fig. 8. Monthly MSP Prediction for onion using RFA.

Table 2. RFA result (Analysis of Fig 8).

Month	Actual Value	Predicted Value	Error
1	1979	1898.49	4.07
2	1797	1678.28	6.61
3	1133	1177.26	3.91
4	1039	1159.81	11.63
5	1516	1517.88	0.12
6	1114	1210.56	8.67
7	1138	1157.61	1.72
8	1168	1167.11	0.08
9	1489	1458.67	2.04
10	1557	1520.25	2.36
11	1087.5	1128.44	3.76
12	1791	1626.46	9.19

From fig. 8 and table 1 following improvements over ANN has been observed

1. Less error in prediction.
2. Peak values are considered.
3. Faster than other algorithms.
4. Lesser memory utilization.
5. Error Percentage - 31%.

IV. Conclusion and Future Scope

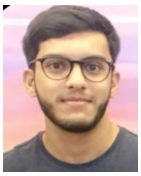
In conclusion, the system is designed to favour the farmers and empower them with information and choice, the system mainly aims to eliminate farmer suicides which is a major issue faced by the country currently. The system will, overtime, suggest farmers which is the best crop to produce, the best time period for cultivating a crop and the best quantity required by the market, ultimately resulting in a profitable yield. The MSP Price is predicted based on multiple factors which acts as a relief factor to the farmer that the production will at least get the projected sum of money and the efforts put in will actually show some good results in terms of money. Thus, bridging the demand supply gap, reducing overproduction and allowing the farmer to get a fair price of the produce at the market.

As a future work, it would be interesting to go to the depth of farming and explore the at most minor details of each and every crop and feed that data into the computer and devise an algorithm that can give us the most accurate data. Furthermore the system will keep track of each and every farmer's land details, pattern of cultivation, choice of fertilizers, seeds and the time taken from the day of sowing to delivery. By the help of machine learning, best suggestions could be given to the farmer for future production based on the needs of the market, season, weather, transportation cost and taking various other factors into consideration.

References

[1] Indian Agriculture and its Problems: https: //www.careeranna.com/articles/indian-agriculture-and-its-problems
[2] Agriculture in India: Information about Indian Agriculture & Its Importance https://www.ibef.org/industry/agriculture-india.aspx
[3] FAOSTAT, 2010 Data, Food & Agriculture Organization of the United Nations. Larry Blume, David A. Easley, Jon M. Kleinberg, and Eva Tardos. Trading networks with price-setting agents.
[4] Tesseract (software) https://en.wikipedia.org/wiki/Tesseract_(software).
[5] Tesseract_OCR https://opensource.google.com/projects/tesseract
[6] The advantages of Google Maps https://www.busymonk.com/advantages-of-google-maps
[7] Google Translate https://en.wikipedia.org/wiki/Google_Translate
[8] Random Forest Machine Learning http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/
[9] Niklas Donges https://medium.com/p/the-random-forest-algorithm-d457d499ffcd
[10] Random Forest https://en.wikipedia.org/wiki/Random_forest
[11] Intelligent Computing theories and Application: 14th International Conference, ICIC 2018, Wuhan, China, August 15-18, 2018, Proceedings.
[12] Fig. 1: Image URL https://medium.com/@balaajip/optical-character-recognition-99aba2dad314
[13] Fig 2 - 3: Image URL http://annadata.yogiprints.com
[14] Fig 4: Image URL https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d
[15] Fig 5: Image URL https://en.wikipedia.org/wiki/File:Colored_neural_network.svg

Author's Profile



Dhruva Phulsundar
B.E. Computer Engineering, Datta Meghe College of Engineering, Airoli, Navi Mumbai, India.
email id: dhruv.phulsundar@gmail.com

Mitalee Minde
B.E Computer Engineering, Datta Meghe College of Engineering, Airoli, Navi Mumbai, India.
email id: mitaleeminde@gmail.com

Saurabh Shingte
B.E Computer Engineering, Datta Meghe College of Engineering, Airoli, Navi Mumbai, India.
email id: saurabh.shingte@gmail.com

Shubham Gujare
B.E Computer Engineering, Datta Meghe College of Engineering, Airoli, Navi Mumbai, India.
email id: shubh3010@gmail.com

Dr. Shubhangi Vaikole
PhD Computer Engineering, Associate. Prof, Datta Meghe College of Engineering, Airoli, Navi Mumbai, India.
email id: slv.cm.dmce@gmail.com