FarmEasy: An Intelligent Platform to Empower Crops Prediction and Crops Marketing

Md Ishak
Computer Science and Engineering
Military Institute of Science and Technology
Dhaka, Bangladesh
ishakahmed4040@gmail.com

Md Shahidur Rahaman
Computer Science
Baylor University
Texas, United States
shahidur.rahaman005@gmail.com

Tahasin Mahmud
Computer Science and Engineering
Military Institute of Science and Technology
Dhaka, Bangladesh
tahasin181962@gmail.com

2021 13th International Conference on Information & Communication Technology and System (ICTS) | 978-1-6654-4059-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICTS52701.2021.9608436

***Abstract*—Farmers' contribution to the economy and national GDP is ineffable. Even though there has been a significant technological advancement in the field of agricultural crops production and management, farmers in developing countries still follow the traditional methods of farming which at many times leads them a loss. Moreover, they don't know the correct market value of their crops, and distributors befool them with the price and value. On the other hand, in times of price hikes and crisis, due to the proper channel government failed to buy crops from them. This paper aims to analyze the primitive approach of cultivation and develop a model for crop prediction using machine learning and provide a model for proper crops management after production. We propose an intelligent system that can predict the best possible crops only by providing the present location of a farmer, the overall guideline from soil preparation to crop yielding, and the systemic approach of crops marketing from farmer to consumer. We used Random Forest Regression, Support Vector Regression and Voting Regression techniques for crop yield prediction and used the real-time data of climate, weather, and soil for the specific region. On the other hand, the market monitoring system will help for proper pricing of the crops and provide transparency for all the stakeholders related to crop marketing where they can buy and sell their products utilizing our system.**

***Keywords*—Crops Prediction, Crops Yield, Machine Learning, Crops Pricing, Smart Farming**

I. Introduction

Agriculture is regarded as one of the nation's most active economic forces. It accounts for 6.4 percent of total global economic output [1]. As the world undergoes a massive transition in the realm of technology, many countries have used smart farming [2] to grow and manage crops. Though smart farming is not a new concept, most nations have quite a minimal level of technological adaptation when it comes to growing crops. Due to recent climate change, farmers increasingly find it challenging to sustain both crop yield and marketing with effective post-harvest care. It is critical for them to understand which crops to choose in certain weather circumstances, especially when it comes to producing crops. Most of the time, they apply fertilizers based on their own assumptions and perpetuate the tradition of growing such crops, even though they do not necessarily understand the soil, climate, or other aspects that affect cultivation. Approximately 1.5 million hectares of land in Bangladesh are effectively flooded every year [3]. Consequently, there are some areas where these cyclones are very severe, and the climate, particularly during droughts, has a significant impact. Drought caused the loss of 2.1 million metric tons of paddy between 1973 and 1987 [4]. During the cultivation phase, various basic techniques must be followed, such as soil preparation, sowing, fertilization, irrigation, harvesting, and preservation. Though they typically follow these, there are several factors that farmers must adapt to when using a scientific method. Considering that 54 percent of the population are highly involved in farming, it is highly improbable for them to expect adequate assistance via digital media.

Producing crops, on the other hand, is not everything, and it is also essential for the farmer to get the optimal rate for his products. Farmers typically sell their goods to local distributors, and many times they may not receive the best price owing to a lack of market knowledge. Finally, retailers purchased from wholesalers, and consumers received the crops that were accessible to them. For these phases there are shifting of prices where many of the factors like processing, labor cost, transportation are included need to be considered. However, many of the instances where price shifting has been imprecise in the stages where the products and goods are being sold. Assume that at some point, a group of distributors forms a syndicate and sets a price that is higher than the purchase price. Many times, the government also unable to do much due to a lack of appropriate oversight of general prices in local or city markets. As a result, there should be a path or flow that allows the government to adequately oversee each market and agricultural selling phase without experiencing obstacles. The objective of this research is to propose a smart system that can guide farmers and monitor in both production and market stages so that farmers can grow richer crops and earn the greatest potential price for themselves. The paper is divided into several sections.

In section II, the objective of the research is focused. Section III gives an idea of related works relevant to this research. In Section IV, the methodology is discussed which involves data preparation, data description, data synthesis, and workflow of the proposed system. Section V presents the experimental results of the study. In section VI, the features of our smart farming system is described. Finally, Section VII concludes the paper with limitations and future works.

II. Objectives

Following the review of Bangladesh's agricultural background, several concerns were identified. First, since farmers in Bangladesh are inexperienced or ignorant of 'smart farming,' they continue to use primitive approaches in the farming process. Second, the framers are denied the right to a fair compensation.

Therefore, this study's objective is to propose a software framework to guide the farmers in maximizing the production output and to organize an e-marketing system that will provide the farmers fair selling price of their crop. The objective can be archived for the following secondary purposes:

- To recommend farmers on which crops are most suited for cultivation in various areas
- To estimate crop yield in different areas of Bangladesh based on weather and soil parameters
- Develop machine-learning models to predict crop yield effectively
- To present sufficient weekly guidelines to farmers for cultivating lands, transplanting seeds and harvesting crops so that production output can be maximized.
- By evaluating the overall cost of production and the existing market value of the crop, determine the optimal price of farmers' crop
- To introduce a farmer-friendly market monitoring system that will ensure that farmers are paid a fair pay

III. Related Works

This section reviews the existing literature on crop yield prediction. The relevant studies were searched in all major scholarly databases and many studies were found. An overview of the articles that have been evaluated is presented below.

Most of the related works for predicting crop yield was carried out by using soil and weather data. For example, Priya et al. [5] researched to predict crop gain with the aid of Random Forest (RF). Historical data (e.g., weather parameter, soil parameter and crop yield) of India during the period of 1997 - 2013 was used where the RF model showed satisfactory accuracy. Similarly, Shakoor et al. [6] investigated the potential of Decision Tree (DT) and K-Nearest Neighbors Regression (KNNR) for estimating crop yield in particular areas. Six major crops in different regions of Bangladesh were considered, and the result showed that DT had a lesser error percentage than KNNR.

Some studies recommended suitable fertilizers for soil as well as predicted crop yield. Bondre and Mahagaonkar [7] made yield prediction with RF and SVM. After soil classification and crop prediction, suitable fertilizer was suggested for each crop. For crop prediction, SVM showed 99.47% accuracy, and RF showed 97.4% accuracy on average. In another study, Bhanumathi et al. [8] used RF and Backpropagation algorithm for the same purposes where the RF model had lesser prediction error rate. Vishwakarma et al. [9] used DT and SVM for predicting appropriate crops based on weather parameters. The proposed system also provided suggestion on fertilizers and seed required for cultivation. A software application was implemented based on the proposed methodology.

Many of the studies used Support Vector Machine (SVM) or Support Vector Regression (SVR) for predicting crop yield. Hossain el al. [10] proposed a weather-based prediction system for rice yield (Aus, Amon, and Boro) for different regions of Bangladesh with Support Vector Regression (SVR). Gandhi et al. [11] used SVM based Sequential Minimal Optimization (SMO) for estimating rice yield which showed about 78.8% accuracy.

Rahman el al. [12] presented a technique for classifying soil series based on soil characteristics. The proposed system also recommended suitable crop for specific soil series. Several machine learning techniques were used and they are KNN, SVM and Bagged Trees while SVM showed highest accuracy. Again, Ahamed et al. [13] used different weather (e.g., rainfall, humidity, temperature), biotic (e.g., max pH, min pH), and area central parameters to estimate the yield of different cops of Bangladesh.

IV. Methodology

First, the objectives of this study were determined. Then, after reviewing related articles, two different machine learning algorithms were selected for developing prediction models. The methodology followed throughout this research is outlined in Fig. 1.

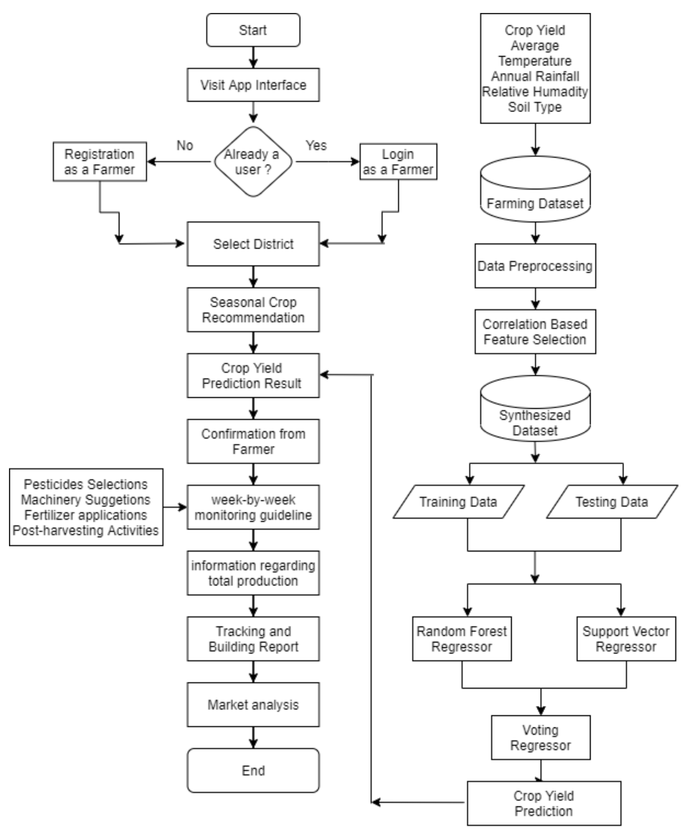

Fig. 1. Methodology of the proposed model

A. Data Preparation

The data collection was put up using information from 'Year Book of Agricultural Statistics of Bangladesh'. The year books of 2015, 2016, 2017, 2018 and 2019 were used to gather soil, weather and crop yield data of 64 districts of Bangladesh during 2013-2019. Seasonal weather data was calculated accordingly. Seasonal weather parameters were obtained by averaging the weather data of four months for respective seasons. For instance, summer season records were obtained from the weather data during March – June, for the rainy season July – October, and finally for winter season weather data during November – February was used. 19 soils unit types were considered to represent soil characteristics of each districts. If a district contains occurrences with a soil unit type, the soil unit type column was set to 'true'; otherwise, the field was set to 'false'. Crop yield for each districts was converted into metric ton/hectare unit.

B. Data Description

To implement the machine learning model, a data set of distinctive features were used. The dataset had 384 instances, each representing weather parameters, soil characteristics and crop yield of 64 districts of Bangladesh. The dataset consists of 39 features, including district; year; relative humidity in summer, rainy and winter; total rainfall in summer, rainy and winter; monthly average minimum temperature in summer, rainy and winter; and monthly average maximum temperature in summer, rainy and winter. This data set also includes the yield of cereal (e.g., Aus, Aman, Boro, wheat, and maize) and pulse (Lentil) crop. Besides these features, 19 soil unit type parameters were used for 64 districts that mark these areas' soil characteristics. After collecting the dataset, it was prepossessed before training the model.

C. Data Synthesis

At this stage, data were preprocessed for preparing it for training machine learning models. At first, null values were replaced by applying suitable replacement techniques where a particular null field was replaced by the mean. Second, the dataset was examined thoroughly to find outliers, and they were removed. Finally, the relation among the independent variables was investigated for detecting multicollinearity occurrences. Multicollinearity weakens the performance of the regression model; hence to solve this problem, one or more variables showing a high correlation was removed as this reduces collinearity among independent variables.

D. Machine Learning Algorithm for Prediction

Dataset was split randomly where 80% of the data was used to train the model, and 20% of the data was used as test data to evaluate the machine learning models. Weather parameters were chosen based on the sowing season of the crop. For example, the sowing season of Aus is March to mid-April. Therefore for training models, weather parameters of the summer season were chosen for Aus. To develop these models, an open sources machine learning library scikit-learn [14] was used. The performance of these algorithms is analyzed in the next section.

Machine learning (ML) is an application of artificial intelligence, a technique of finding the undiscovered pattern in data, learning from data, and making a decision with minimal human intervention. Typical application of machine learning include image recognition [15], medical diagnosis [16], web mining [17], autonomous car, stock market analysis [18], language translation, product recommendation [19], sentiment analysis [20] and the list goes on. The use of ML is becoming more common as the volume of structured and unstructured data grows.

In this study, supervised regression rather than classification algorithms was used as the ML models had to predict continuous values of crop yield. Supervised learning trains the model with the labeled dataset. After reviewing the related works, two different machine learning algorithms Random Forest Regression (RFR) and Support Vector Regression (SVR), were chosen for developing the prediction model. Then Voting Regression (VR) was proposed using RFR and SVR as the base model.

1) Random Forest Regression: Random forest regression (RFR) is a supervised learning algorithm for regression using the ensemble learning method. Random forest is an ensemble of decision trees, while each tree makes its own prediction. The final output is the mean of all the base classifiers. The averaging enables RFR to perform better than decision tree regression.

2) Support Vector Regression: The support vector regression (SVR) algorithm is part of the support vector machine family (SVM). SVM is widely for classification problems in machine learning. The goal of SVM is to create a decision

TABLE I
PERFORMANCE OF THE SELECTED ALGORITHMS

	Random Forest Regression		**Support Vector Regression**		**Voting Regression**	
	R-Squared value	Root Mean Squared Error	R-Squared value	Root Mean Squared Error	R-Squared value	Root Mean Squared Error
Aus Rice	74.5%	23.4%	66.8%	26.6%	75.4%	22.9%
Aman Rice	79%	21.9%	81.2%	20.7%	83.2%	19.6%
Boro Rice	78.7%	24.3%	89.9%	16.7%	87%	19%
Wheat	78%	42%	88.7%	32%	87.4%	34.4%
Lentil	65.5%	16.8%	80%	12.6%	81.2%	11.3%
Maize	75.2%	21.4%	79.5%	21.7%	82.6%	20.3%
Average	75.15%	25%	81%	21.7%	82.8%	21.3%

boundary or hyperplane that separates data points with different class labels. Support vectors are the data points nearest to the hyperplane. If these data points are removed, then the position of the decision boundary or hyperplane will change. Instead of producing a decision boundary, support vector regression (SVR) uses training data to derive a function that predicts numerical values.

3) Voting Regression: A voting regression (VR) is an ensemble machine learning model that combines the predictions from multiple base models. VR calculates the average of prediction from the base models. The proposed VR algorithm is outlined in Algorithm 1.

Algorithm 1: Proposed Voting Regression Algorithm for Crop Yield

---

Preproceess the given data as described;
Train the machine with `DATA` ;
Test the trained machine ;
`pred_RFR` = crop yield from Model with Random Forest Regression ;
`pred_SVR` = crop yield from Model with Support Vector Regression ;
pred_final Will be result of Voting Regressor
    `pred_final` $\leftarrow$ *Average*(`pred_RFR`,`pred_SVR`)

V. Experimental Result

Based on the performance of the algorithms in relevant previous studies, two different machine learning models were chosen. These two models were combined to develop an ensemble regression, namely voting regression. The r-squared value and root mean squared error was used to evaluate the regression models' performance. The rest of this section analyzes the performance of developed machine learning models.

A. Random Forest Regression

Performance of RFR is displayed in Table I, Fig. 2 and Fig. 3. It's clear that the other two regression models outperformed RFR. The highest coefficient of determination or the r-squared value was obtained for predicting the yield of Aman rice (79%). The average r-squared value for the RFR algorithm is 75.15%, and the average root mean squared error is 25%.

B. Support Vector Regression

SVR gave the best result for predicting the yield of Boro and Wheat. The average r-squared value for the RFR algorithm is 81%, and the average root mean squared error is 21.7%.

C. Voting Regression

The proposed VR evaluates the mean of crop yield prediction from RFR and SVR models. VR outperforms the other two regression models in terms of r-squared value, and root mean squared error. VR game best prediction result for predicting the yield of aus, aman, lentil, and maize. The average r-squared value and root mean squared error for predicting the yield of six crops are better than the average

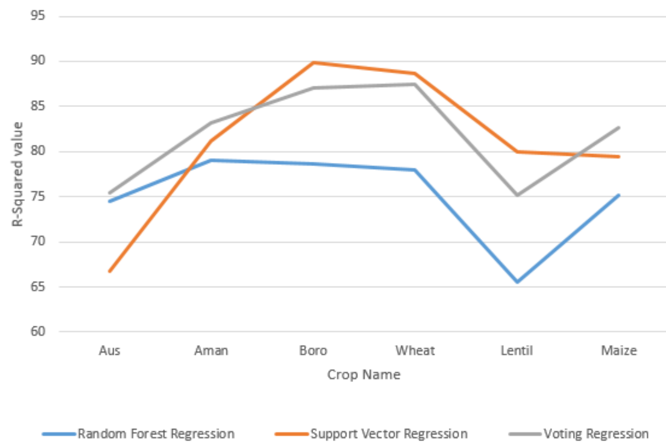

Fig. 2. Performance (r-squared value) of the selected algorithms

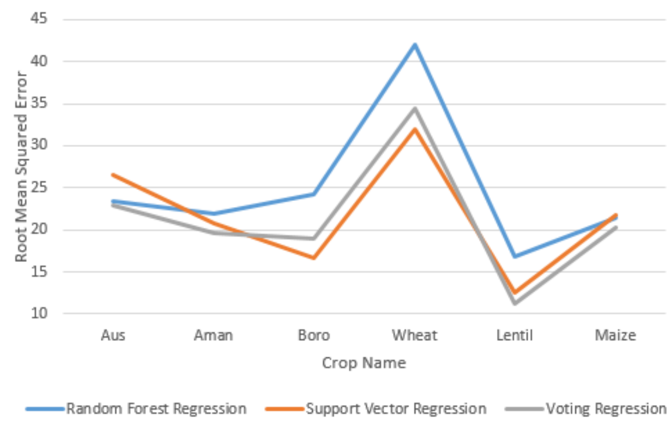

Fig. 3. Performance (root mean squared error) of the selected algorithms

of the other two algorithms (82.8% and 21.3%, respectively). Hence, the proposed VR model is the best among the three algorithms for predicting crop yield.

VI. Recommendation For Farming

In this research, The crop production procedure is outlined in Fig. 1 and market monitoring system is outlined in Fig. 4. So, for implementing our research idea, a mobile application is proposed that includes the following characteristics.

1) Stakeholder Registration: To join up, all the farmers, distributors, crop processors, wholesalers, retailers, and consumers will need to register by providing the required information. Then a concerned user who registered as a stakeholder might provide some essential data for successful login. After successful completion of the procedure, a district or place can be picked according to the stakeholders' preference.

2) Crop Recommendation and Yield Prediction:

- District Selection of the farmers
- Seasonal crop recommendation according to the chosen district
- Assisting farmers in selecting crops effectively

- Provide crop yield prediction through app and text message

3) Week by week guideline and proper monitoring system:

- Calculating amount of seed required for cultivation
- Land preparation and Seed sowing guideline
- Fertilizer measurement Pesticides selection guideline
- Irrigation time management guideline
- Weeding procedure and Harvesting technique guideline

4) Agriculture Crop Security:

- Storing the information of entire yield of the farmers
- Generating report of the total production of the territory
- Proper crop storage facilities
- Ensures crop quality and packaging by distributor and trader

5) Crop Price Estimation: After the successful production, the post-harvesting procedure will be contemplated. After scrutinizing the current market value, overall harvest cost, and previous price estimation, the app will offer an optimum price ratio. The farmer will check through cost estimation options, and the farmer will pick a preferable price rate for selling the crop. Then when a trader finished storing, he set an estimated price based on the farmers' choice for wholesalers. When a registered user login as a wholesaler, he can only pick his desirable price from all the prices submitted by the farmers in that territory and can directly contact that trader. Then a user interested in retailing or exporting will choose a price given by the wholesaler and can directly contact the wholesaler. After that, a consumer can purchase a product through the app or megastore at an affordable price.

6) Price hike solution:

- Communication system is created at farmer level
- Facility of selling crops at fair rates
- Government can identify dishonest businessman
- Ensuring the comfort of common people
- Eliminating the price-hikes of essential commodities

7) Other Features:

- Excellent platform for emerging entrepreneurs
- Offer an entrepreneur to buy agricultural foods at wholesale price
- Planning to establish a farmer-consumer market for more profitable prices

VII. LIMITATIONS AND FUTURE WORKS

The data set was prepared by collecting the information from the 'Year Book of Agricultural Statistics of Bangladesh'. But if we conducted a field survey to gather the data, then the accuracy of the data could be more significant. We have utilized in total 8 years of data (2013-2020). The study comprises of six main crops because of data unavailability of additional crops. Another 10 to 15 additional years of data and 6 to 7 crops data could be useful to obtain a more accurate prediction.

Here, a mobile application model is proposed, hence we are planning to build the mobile application for all types of platforms and also planning to execute the web application part. We also attempt to integrate image processing as our model can identify crop disease and may provide pesticide prediction based on that disease. We are also trying to add additional features such as price prediction and a loan management system.

VIII. CONCLUSION

This research indicates that the precise prediction of various kinds of crop yields across different districts may assist many farmers and others the same. The proposed system tried to provide a prediction on how a properly coordinated machine learning model and good market monitoring might alter our country's agricultural image. Our administration has successfully adopted such endless outstanding initiatives in the agriculture sector. Therefore, this consequence of these efforts will be beneficial to new entrepreneurs, consumers, and farmers. Moreover, the proposed software solution may bring about advantages and establish cutting-edge cultivating techniques that can operate on our economy and facilitate us with standing apart as an advanced country. Finally, Our mobile application will be a great solution to provide a digital agrarian framework for optimum crop selection, crop yield prediction, and market stability.

REFERENCES

[1] "List of countries by gdp sector composition," https://statisticstimes.com/economy/countries-by-gdp-sector-composition.php, accessed: 2021-07-12.

[2] P. Tripicchio, M. Satler, G. Dabisias, E. Ruffaldi, and C. A. Avizzano, "Towards smart farming and sustainable agriculture with drones," in *2015 International Conference on Intelligent Environments*, 2015, pp. 140–143.

[3] "Climate change and it's impact in bangladesh," https://ejfoundation.org/reports/climate-displacement-in-bangladesh, accessed: 2021-07-12.

[4] M. N. Islam and A. van Amstel, *Bangladesh I: climate change impacts, mitigation and adaptation in developing countries*. Springer, 2019.

[5] P. Priya, U. Muthaiah, and M. Balamurugan, "Predicting yield of the crop using machine learning algorithm," *International Journal of Engineering Sciences & Research Technology*, vol. 7, no. 1, pp. 1–7, 2018.

[6] M. T. Shakoor, K. Rahman, S. N. Rayta, and A. Chakrabarty, "Agricultural production output prediction using supervised machine learning techniques," in *2017 1st international conference on next generation computing applications (NextComp)*. IEEE, 2017, pp. 182–187.

[7] D. A. Bondre and S. Mahagaonkar, "Prediction of crop yield and fertilizer recommendation using machine learning algorithms," *International Journal of Engineering Applied Sciences and Technology*, vol. 4, no. 5, pp. 371–376, 2019.

[8] S. Bhanumathi, M. Vineeth, and N. Rohit, "Crop yield prediction and efficient use of fertilizers," in *2019 International Conference on Communication and Signal Processing (ICCSP)*. IEEE, 2019, pp. 0769–0773.

[9] N. K. Dhanush Vishwakarma and M. N, "Crop prediction using machine learning approaches," *International Journal of Engineering Research and*, vol. V9, no. 08, Aug. 2020. [Online]. Available: https://doi.org/10.17577/ijertv9is080029

[10] M. A. Hossain, M. N. Uddin, M. A. Hossain, and Y. M. Jang, "Predicting rice yield for bangladesh by exploiting weather conditions," in *2017 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2017, pp. 589–594.

[11] N. Gandhi, L. J. Armstrong, O. Petkar, and A. K. Tripathy, "Rice crop yield prediction in india using support vector machines," in *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE, 2016, pp. 1–5.

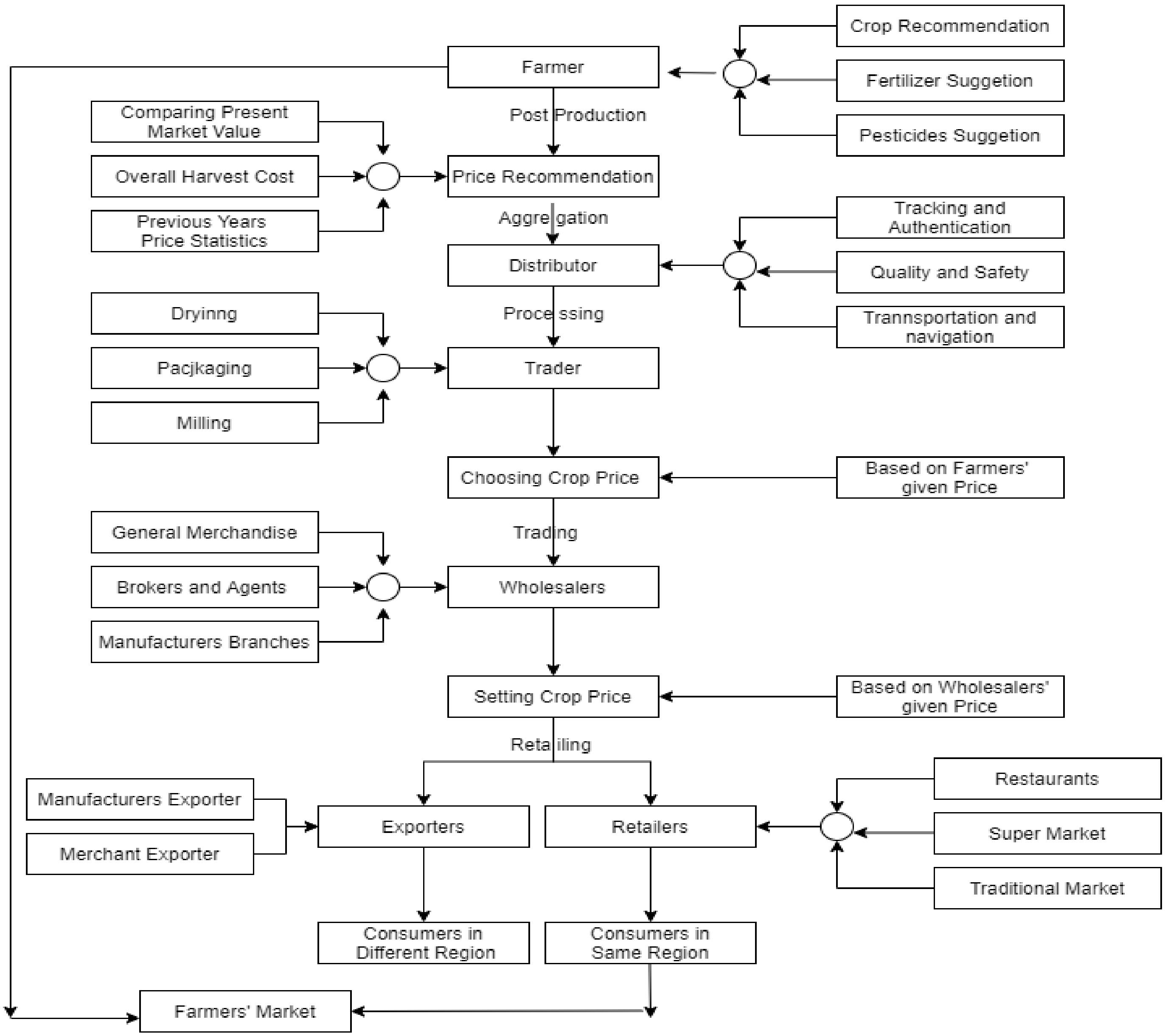

Fig. 4. Proposed market monitoring system

[12] S. A. Z. Rahman, K. C. Mitra, and S. M. Islam, "Soil classification using machine learning methods and crop suggestion based on soil series," in *2018 21st International Conference of Computer and Information Technology (ICCIT)*. IEEE, 2018, pp. 1–4.

[13] A. M. S. Ahamed, N. T. Mahmood, N. Hossain, M. T. Kabir, K. Das, F. Rahman, and R. M. Rahman, "Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in bangladesh," in *2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. IEEE, 2015, pp. 1–6.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[15] M. Wu and L. Chen, "Image recognition based on deep learning," in *2015 Chinese Automation Congress (CAC)*. IEEE, 2015, pp. 542–546.

[16] K. R. Foster, R. Koprowski, and J. D. Skufca, "Machine learning, medical diagnosis, and biomedical engineering research-commentary," *Biomedical engineering online*, vol. 13, no. 1, pp. 1–9, 2014.

[17] H. Chen and M. Chau, "Web mining: Machine learning for web applications," *Annual review of information science and technology*, vol. 38, no. 1, pp. 289–329, 2004.

[18] R. Choudhry and K. Garg, "A hybrid machine learning system for stock market forecasting," *World Academy of Science, Engineering and Technology*, vol. 39, no. 3, pp. 315–318, 2008.

[19] H. Tuinhof, C. Pirker, and M. Haltmeier, "Image-based fashion product recommendation with deep learning," in *International Conference on Machine Learning, Optimization, and Data Science*. Springer, 2018, pp. 472–481.

[20] M. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," in *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*. IEEE, 2013, pp. 1–5.