International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Empowering Agriculture: A Machine Learning-Based Decision Support System for Crop Selection and Profitability Analysis

T. Jalaja[1], Dr. T. Adilakshmi[2], Meka Dheeraj Reddy[3], K. Sai Kamal[4]

Submitted: 25/01/2024 **Revised**: 03/03/2024 **Accepted**: 11/03/2024

Abstract: Modern agriculture is increasingly reliant on technological advancements to optimize productivity and decision-making processes. This paper introduces a sophisticated web application designed to help farmers in selecting the appropriate crops for their soil, considering soil composition, weather conditions, market trends, and financial factors. The application integrates various machine learning methodologies to deliver personalized recommendations tailored to each farmer's specific circumstances. Initially, farmers provide details of soil mineral percentages, focusing on Nitrogen (N), Phosphorus (P), and Potassium (K) to the application. Subsequently, real-time weather data is retrieved from a Weather API to evaluate environmental suitability. An ensemble machine learning model, comprising Support Vector Classifier(SVC), K-Nearest Neighbours (KNN), Gaussian Naive Bayes Classifier, and Decision tree, is deployed to identify the top three crops best suited to the soil. To refine decision-making further, a Stacking Ensemble technique is employed, leveraging Regression algorithms such as Random Forest and XGBoost Regressors to forecast the market selling price of each crop based on historical data. This empowers farmers to assess both suitability and profitability concurrently. Additionally, the application provides detailed insights into cultivation costs per acre, aiding farmers in financial planning and risk management. Furthermore, instructional videos on crop cultivation practices are integrated to facilitate knowledge transfer and skill enhancement among farmers. Through elucidating the architecture, functionality, and performance metrics of the web application, this paper underscores it's potential to revolutionize agricultural decision-making processes, enhance farm productivity, and ultimately contribute to sustainable rural livelihoods.

Keywords: *K-Nearest Neighbours (KNN), Support Vector Classifier(SVC), Gaussian Naive Bayes Classifier, Decision Tree, Ensemble Classifier, Stacking Ensemble Technique, XGBoost, Profitability.*

1. Introduction

Agriculture, as the backbone of human civilization, constantly evolves to meet the challenges of a growing global population, changing climate patterns, and economic fluctuations. In this dynamic landscape, the integration of technology has become imperative to enhance productivity, optimize resource utilization, and facilitate informed decision-making processes for farmers. In response to these demands, we present a novel approach: a sophisticated web application tailored to help farmers in selecting the most appropriate crops for a soil. This paper outlines the development and implementation of a comprehensive agricultural decision support system that harnesses the power of Machine Learning, real-time Data analytics, and instructional resources.

[1] *Professor & HOD - Computer Science and Engineering Department, Vasavi College of Engineering, Hyderabad, India*
[2] *Assistant Professor- Computer Science and Engineering Department, Vasavi College of Engineering, Hyderabad, India.*
[3,4] *Computer Science and Engineering Department, Vasavi College of Engineering, Hyderabad, India*
* *Corresponding Author Email t_adilakshmi@staff.vce.ac.in*
jalaja.t@staff.vce.ac.in saikamalkonduru@gmail.com
dheerajreddymeka@gmail.com

The Web application's primary objective is, to empower farmers with actionable insights derived from an amalgamation of soil composition analysis, weather forecasting, market trends, and financial considerations. The significance of this project lies in its ability to address multiple dimensions of agricultural decision-making simultaneously. By integrating diverse datasets and employing advanced machine learning methodologies, the application provides personalized recommendations tailored to each farmer's unique circumstances. Through this approach, farmers gain access to a holistic decision-making tool that considers not only soil compatibility but also profitability and risk mitigation strategies.

The foundation of the web application rests on two key pillars: soil analysis and predictive modeling. Initially, farmers give details of essential soil mineral percentages, focusing on Nitrogen (N), Phosphorus (P), and Potassium (K) in the application UI. Leveraging this information alongside real-time weather data retrieved from a Weather API, the application evaluates environmental suitability for various crops. An ensemble machine learning model, comprising K-Nearest Neighbors (KNN), Gaussian Naive Bayes Classifier, Decision Tree, and Support Vector Classifier, is then deployed to identify the top three crops best suited to the soil. To refine decision-making further and enable farmers to assess both suitability and profitability concurrently, a Stacking Ensemble technique is employed.

This technique leverages Regression algorithms such as Random Forest and XGBoost regressors to forecast the market selling price of each crop based on historical data. Additionally, the application provides detailed insights into cultivation costs per acre, aiding farmers in financial planning and risk management.

2. Related Work

[1] "Crop Price Prediction System using Machine Learning Algorithms" emphasizes the critical nature of price prediction in agriculture, particularly in forecasting crop prices for the future rotation based on real data. The paper underscores the importance of thorough data analysis, cleansing, and exploratory data analysis (EDA) to comprehend the various parameters within the dataset. Employing diverse data mining techniques, the study aimed to construct an accurate model for precise price prediction. Various Machine learning algorithms, including Decision Trees, Linear regression, and XGBoost were employed for this purpose. Notably, XGBoost demonstrated superior performance among these algorithms in predicting crop prices, showcasing its effectiveness in the agricultural price prediction system.

[2] "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector" suggested a way to predict the crop yield from the historical data that includes factors like Ph, rainfall, name of crop, humidity, and temperature. They have used to algorithms random forest, decision tree and logistic regression, of which Random Forest has provided more accurate result.

The paper entitled [3] "Stacking Strong Ensembles of Classifiers" has clearly explained how the stacking technique can be leveraged to combine the predictions of multiple prediction models and generate a meta model that can be used trained and can be used to predict the results. It has also discussed the architecture of a stacking ensemble model which consisted of different components in order to build an effective model that one needs to take care of.

3. Dataset

We have used two datasets, one for Crop Prediction and one for Crop Market Selling Price Estimation.

3.1 Crop Prediction Dataset

We chose crop prediction dataset from Kaggle which contains fields like Nitrogen(N%), Phosphorous(P), Potassium(K) percentages, Rainfall, Humidity, Temperature, Crop.

Temperature (°C)	Humidity (%)	Rainfall (mm)	N (ratio)	P (ratio)	K (ratio)	Crop
20.88	82	202.94	0.51	0.24	0.25	Rice
21.77	80.32	226.66	0.46	0.32	0.22	Rice
23	82.32	263.96	0.38	0.35	0.28	Rice
26.49	80.16	242.86	0.5	0.23	0.27	Rice
20.13	81.6	262.72	0.48	0.26	0.26	Rice
23.06	83.37	251.05	0.47	0.25	0.28	Rice
22.71	82.64	271.32	0.43	0.34	0.23	Rice
20.28	82.89	241.97	0.5	0.28	0.21	Rice
24.52	83.54	230.45	0.49	0.3	0.21	Rice
23.22	83.03	221.21	0.41	0.35	0.23	Rice
26.53	81.42	264.61	0.49	0.29	0.22	Rice
23.98	81.45	250.08	0.51	0.26	0.24	Rice
26.8	80.89	284.44	0.43	0.32	0.24	Rice
24.01	82.06	185.28	0.5	0.3	0.19	Rice
25.67	80.66	209.59	0.52	0.28	0.2	Rice

Figure 1. Snapshot of Dataset.

3.2 Crop Market Price Historical Dataset.

(I) District Attribute –Districts in India where the crop is predominantly grown.

(II) Crop Attribute – This data provides details about different crops that are considered for their profit price estimation.

(III) Month Attribute– This data is considered to estimate the profit on a monthly basis for more clarity.

(IV) Crop Profit Price- This column provides insight about the profit obtained per quintal of that respective crop.

Number of records in the dataset are 30000 rows and 4 columns. The Dataset is gathered primarily from two sources: Home | Open Government Data (OGD) Platform India and Socio-Economic Statistics India, Statistical Data Figures Year-Wise (indiastat.com).

District	Crop	Price Date	Crop Profit (Rs per quintal)
Ariyalur	Banana	Jan-15	1830
Ariyalur	Banana	Jan-15	1820
Ariyalur	Banana	Feb-15	1730
Ariyalur	Banana	Mar-15	1780
Ariyalur	Banana	Apr-15	1900
Ariyalur	Banana	May-15	1850
Ariyalur	Banana	Jun-15	1845
Ariyalur	Banana	Jun-15	1875
Ariyalur	Banana	Jul-15	1850
Ariyalur	Banana	Aug-15	1890
Ariyalur	Banana	Sep-15	1710
Ariyalur	Banana	Oct-15	1750
Ariyalur	Banana	Oct-15	1715
Ariyalur	Banana	Nov-15	1725
Ariyalur	Banana	Dec-15	1730

Figure 2. Snapshot of Dataset.

Both the datasets contain details about 10 crops - Banana, Bengal gram, Black gram, Coconut, Coffee, Cotton, Green gram, Maize, Red Gram and Rice. The crop profit price information is gathered monthly for every year for the period of 2015-2022 so as to predict with good accuracy using regression algorithm.

4. Proposed Methodology

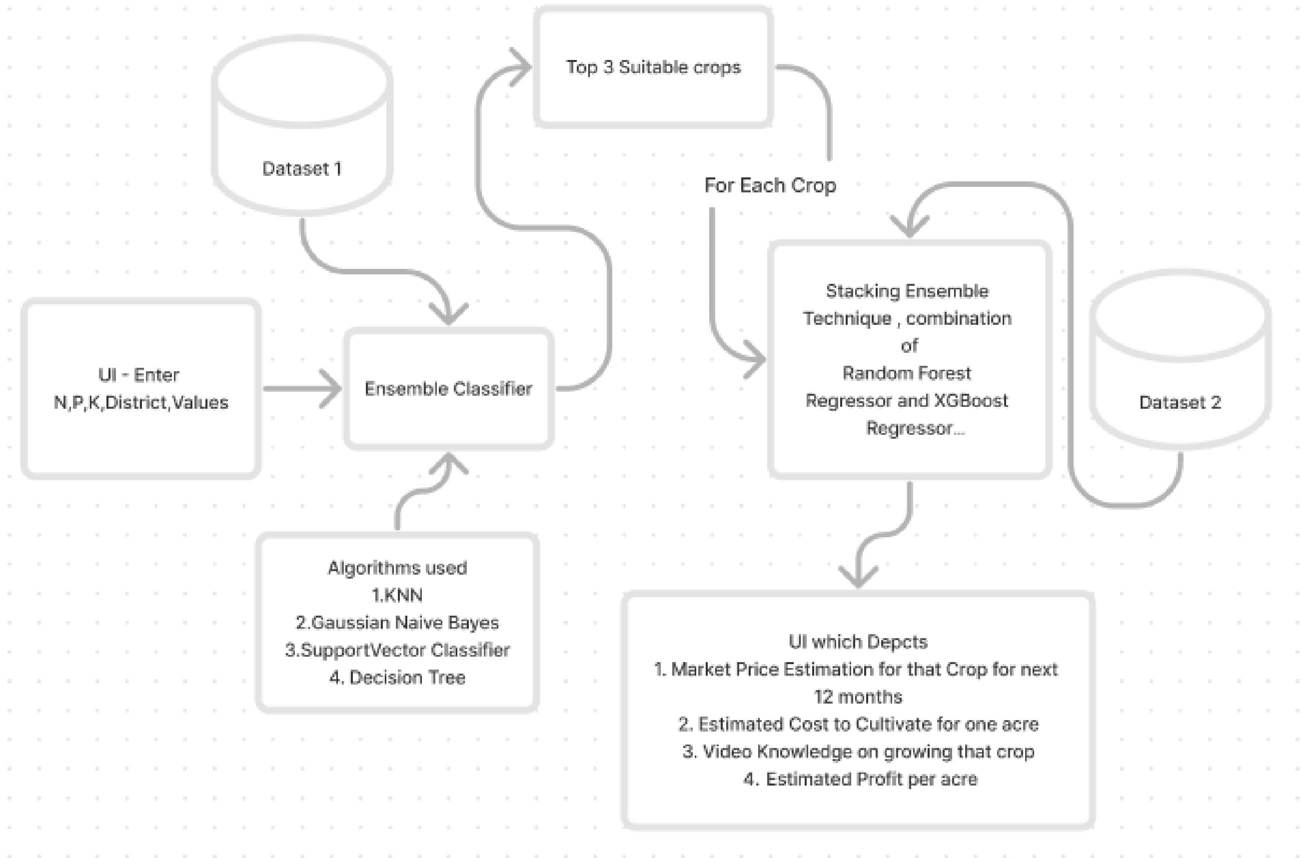


Figure 3 Architecture of the Proposed Solution.

Our proposed methodology outlines a comprehensive approach to crop prediction and crop market price estimation, using advanced machine learning techniques and a user-centric interface. We have divided the methodology into three modules as following:

Module 1: Data Preprocessing

Our Crop prediction dataset contains 2301 rows and 6 different attributes for 10 crops respectively. In Crop prediction dataset we have used Simple Imputer to handle the missing values. Crop Market Price Estimation Dataset contains 30000 rows and 4 columns, it contains past 6 years data for each crop over 12 months of each year. In this dataset, we have used Simple Imputer to handle Missing values and Label Encoder for categorical attributes and Standard Scaler for normalization of price attribute. We have carefully assessed the dataset while data collection and choose the suitable techniques that provide good quality data. This completes the data preprocessing stage.

Module 2: Ensemble Classifier Model Building using various Algorithms

In the next module, we have utilized prominent machine learning algorithms, including Support Vector Machines (SVM), Decision Tree, K-Nearest Neighbors (KNN), and Random Forest. We have leveraged the strengths of each algorithm to capture diverse patterns within the data.

Basic description of algorithms

1**. K-Nearest Neighbors (KNN)**: KNN stands for K-Nearest Neighbors, a straightforward yet powerful algorithm commonly applied in classification and regression tasks. Its operation involves determining the predominant class or average value within the vicinity of the query point, determined by the 'k' nearest data points. In our specific context, we utilize KNN to match soil compositions by their similarity, enabling us to suggest appropriate crops based on their proximity in the feature space.

2. **Support Vector Machine (SVM):** SVM, or Support Vector Machine, is a robust algorithm utilized for various tasks such as classification and regression. Its functionality involves identifying the optimal hyperplane to differentiate between distinct classes or forecast continuous results. Within our application, SVM plays a crucial role in categorizing soil compositions, thereby improving the precision of crop recommendations tailored to specific conditions.

3. **Gaussian Naive Bayes**: Gaussian Naive Bayes is a probabilistic classifier grounded in Bayes' theorem,

presuming independence among features. It proves valuable in classifying tasks involving continuous input variables. In our scenario, Gaussian Naive Bayes aids in forecasting crop suitability by analyzing soil mineral percentages, factoring in their probabilistic distribution.

4. **Decision Tree**: Decision Tree recursively splits data based on the most significant features. In our context, Decision Trees contribute to disease prediction by creating interpretable rules based on symptoms, aiding in the understanding of the underlying patterns in the datasets.

In addition to it we have implemented an ensemble method, specifically a voting mechanism, to amalgamate predictions from individual models. This ensemble approach aims to enhance predictive accuracy and generalizability by leveraging the collective intelligence of multiple algorithms.

```
nb_clf = GaussianNB(var_smoothing=0.0004)
svm_clf = SVC(C=0.05, kernel='linear',probability=True, random_state=5)
knn_clf = KNeighborsClassifier(n_neighbors=25, p=5)
dt_clf = DecisionTreeClassifier(max_depth=6, min_samples_split=20, max_features=2, random_state=42)

voting_clf = VotingClassifier(
    estimators=[('nb', nb_clf), ('svm', svm_clf), ('knn', knn_clf), ('dt', dt_clf)],
    voting='soft')

voting_clf.fit(X_train, y_train)
```

Fig 5. Ensemble method for model building

For every algorithm we have assessed the performance of the developed system through rigorous evaluation metrics, including precision, accuracy, F1-score and recall. We have also validated the system on diverse datasets to ensure robustness and reliability.

Algorithm	Accuracy %
Support Vector Classifier	94.32
Gaussian Naive Bayes	92.19
Decision Tree	95.63
KNN	94.39
Ensemble Technique	98.91

Table 1. Accuracy comparison among various algorithms

To estimate the Market Price of a crop we have leveraged the power two regressors, Random Forest and XGBoost using Stacking Ensemble technique with Linear regression as meta model. This innovative approach has improved the prediction results better than the individual algorithms.

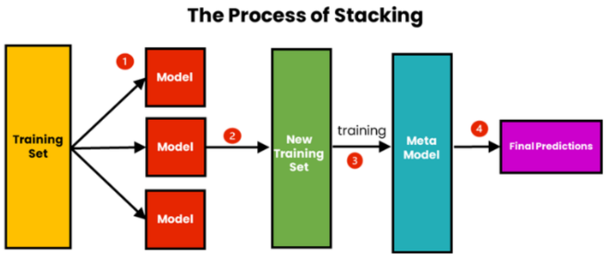


Fig 4. Stacking Ensemble method for price estimation.

Module 3 : Building user friendly interface

In the next module we have developed an intuitive interface that facilitates farmer-friendly soil profile input fields and district dropdown menu. The main objectives in this section include to Prioritize clarity, simplicity, and accessibility to ensure a seamless interaction experience for users with varying levels of technical proficiency. We also enhanced the system's utility by integrating graphs and tables for better visualization. We also embedded a video in each page when a user clicks on a crop to know about more on how to grow that crop. We have gathered the estimated crop cultivation from various credible sources keeping in view various factors like labor, transportation costs etc. We have also included the feature of automatic calculation of profit per acre that a farmer can get, also showing the detailed breakdown of the expenses incurred.

5. Results and Analysis

The results and analysis section of this paper presents a comprehensive evaluation of the performance and efficacy of the developed web application in assisting farmers with crop selection and decision-making processes. Through extensive experimentation and validation, we assess the accuracy, reliability, and practical utility of the application's recommendations.

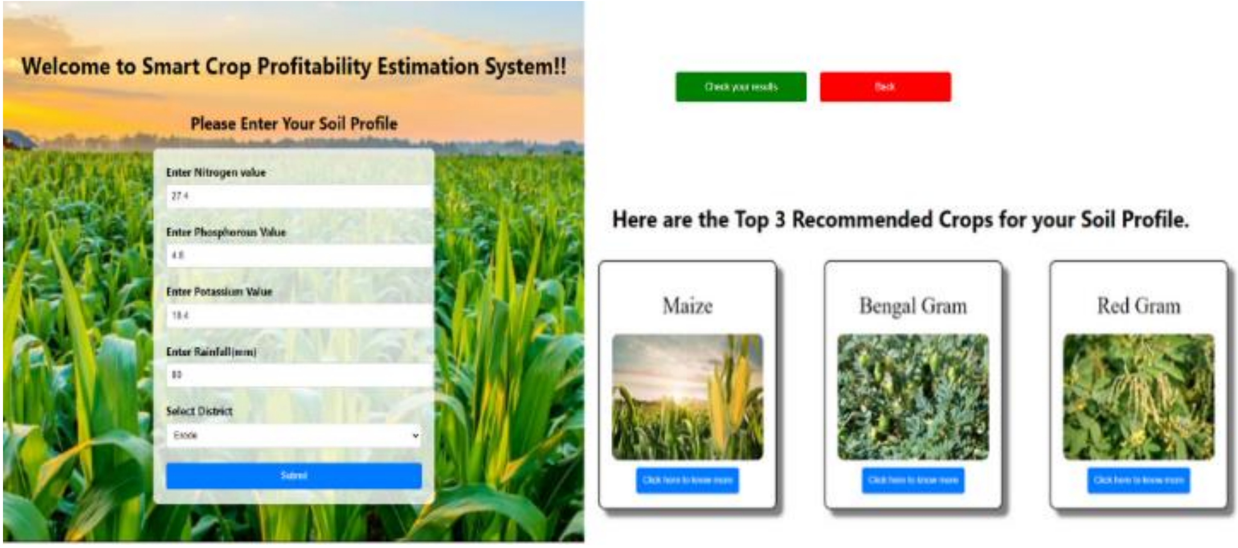
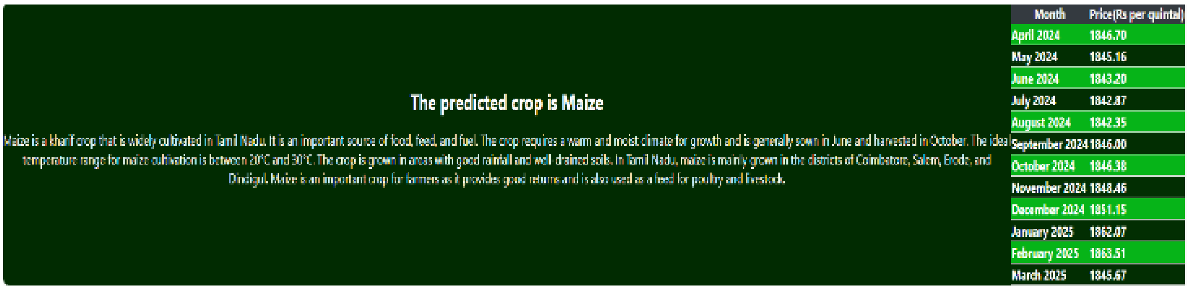


Fig 5. Webpage for farmer to enter input and top 3 crops predicted.



Month	Price(Rs per quintal)
April 2024	1846.70
May 2024	1845.16
June 2024	1843.20
July 2024	1842.87
August 2024	1842.35
September 2024	1846.00
October 2024	1846.38
November 2024	1848.46
December 2024	1851.15
January 2025	1862.07
February 2025	1863.51
March 2025	1845.67

Fig 6(a)

Fig 6(b)

Fig 6. (a) Shows information about crop and estimated prices,(b) Graph that depicts the prices, (c) Video knowledge and cost of 1 acre cultivation and estimated profit that can be obtained.

6. Conclusion and Future scope

In conclusion, the implementation and development of the sophisticated web application presented in this paper represent a significant step forward in revolutionizing agricultural decision-making processes. By leveraging advanced technologies such as machine learning, real-time data analytics, and instructional resources, the application empowers farmers with personalized recommendations tailored to their unique circumstances. Through comprehensive analysis and validation, we have demonstrated the efficacy and practical utility of the application in fostering the decision of farmers with crop selection, resource utilization, and financial planning.

The ensemble machine learning model, comprising Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Gaussian Naive Bayes, and Decision Tree algorithms, provides accurate and reliable crop recommendations based on soil composition and environmental factors. Additionally, the Stacking Ensemble technique enables farmers to assess the profitability of recommended crops by predicting market selling prices, thereby facilitating informed decision-making regarding crop selection.

Furthermore, the inclusion of detailed insights into cultivation costs per acre and instructional videos on crop cultivation practices enhances the application's usability and value proposition for farmers of all backgrounds and resource levels. By democratizing access to agricultural expertise and facilitating knowledge transfer, the application contributes to the empowerment and sustainability of rural livelihoods.

Furthermore, We can develop a feature which shows a list of nearest procurement centers for a crop so that a farmer can estimate his transportation costs easily. We can suggest the best seeds for a crop from the market and where can he procure them or provision him an online order option. We can add a feature of language translation into webpage so that farmers can read the contents in their preferred language.

Overall, the success of this project underscores the transformative potential of technology in addressing the complex challenges of modern agriculture. By connecting traditional farming methods with advanced technological solutions, the web application has the potential to transform agricultural decision-making processes, improve farm productivity, and encourage sustainable rural development. As we look towards the future, continued innovation and collaboration in the field of agricultural technology will be essential in ensuring a more resilient, efficient, and equitable agricultural sector for generations to come.

Conflicts of interest

The authors declare no conflicts of interest.

References and Footnotes

[1] Pandit Samuel, B.Sahithi, T.Saheli, D.Ramanika, N.Anil Kumar "Crop Price Prediction System using Machine learning Algorithms." Quest Journals Journal of Software Engineering And Simulation, Vol. 06, No. 01, 2020,Pp. 14-20.

[2] Kumar, Y. Jeevan Nagendra, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi. "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector".

[3] "Stacking Strong Ensembles of Classifiers" by Stamat

[4] ios-Aggelos N. Alexandropoulos, Christos K. Aridas,

[5] Sotiris B. Kotsiantis, and Michael N. Vrahatis.

[6] Rohith R, Vishnu R, Kishore A, Deeban Chakkarawarthi, "Crop Price Prediction and Forecasting System using Supervised Machine Learning Algorithms", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 9, Issue 3, March 2020.

[7] R. Nagini, "Agriculture yield prediction using predictive analytic techniques", 2nd International Conference on Contemporary Computing and Informatics, 2016.

[8] S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh.2019. Machine learning approach for forecasting crop yield based on climatic parameters.
[9] Yung-HsinPeng, Chin-Shun Hsu, and Po-Chuang Huang Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets, 2017 IEEE.