Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 4 Issue 2 (2024)

Comparative Analysis of Data Mining Techniques to Enhance the Decision Making in Crop Yield.

[1]**Deepak Sharma**
Research Scholar, Jagannath University, Jaipur
Assistant Professor, Jagannath International Management School, New Delhi.
deepaktech@hotmail.com

[2]**Dr. Deepshikha Aggarwal**
Professor, Jagan Institute of Management Studies, Delhi
[3]Pramod Kumar Pandey
Assistant Professor, Jagannath International Management School, New Delhi

Abstract:

Data mining software has been developed by both commercial and academic organisations, and it employs a range of methodologies. These strategies have been used by a variety of organisations, including industrial, commercial, and academic institutions. Massive data sets can be analysed to discover useful categories and trends, for example, through the use of data mining. The use of a variety of data mining technologies to analyse these agricultural data sets is discussed in detail below. The aim of this study is to develop an early prediction strategy for farmers' cost-benefit analyses. This study suggests a number of models and computational approaches in order to prevent agricultural communities from incurring losses or acquiring debt as a result of their efforts. It is possible to make more effective decisions on farm management and agribusiness activities with the assistance of Agriculture Intelligence, such as determining the best cultivars to plant on their farm, determining the optimal cultivation date, Investment Prioritizing, and Evaluate demand and supply risk, with the assistance of Agriculture Intelligence. After that, you must decide on the level of precision.

Keywords: Data Mining, Crop, Decision Making, CDMA

1. Introduction

The growing population and resulting need for increased agricultural production necessitate the urgent need for improved management of agricultural resources [1], which is a prerequisite for increased agricultural productivity. Agriculture provides a means of subsistence for about two-thirds of India's people, and as a result, agriculture is the economic backbone of the country. It is estimated that just one-third of the planted area is irrigated, resulting in extremely low agricultural output [2, 3]. As a result, as the need for food rises, farmers, agricultural scientists, and the government are attempting to make a greater effort by using strategies that will increase food production levels. Farmers, on the other hand, continue to conduct agriculture-related jobs manually, with just a small number of farmers employing innovative farming methods, tools, and techniques to improve agricultural production.

The uniform input application that is common in traditional agricultural field management ignores not only the concept of geographical and temporal diversity within a crop field, but it also results in environmental contamination and a decline of farm income [4-7]. Researchers, producers, and farmers from all over the world have argued for the importance of site-specific management, often known as precision agriculture, in their operations. Precision agriculture is built on advanced information technology that can quickly and cost-effectively identify spatial heterogeneity within crop fields. This is the foundation of precision agriculture. Furthermore, remote sensing technologies have progressed fast in recent years and have become excellent tools for site-specific management in crop protection and crop production, among other applications.

2. Background

In data analysis, classification can be used to generate models characterising key data classes or to forecast future trends. Discrete class labels can be predicted from a model's ability to learn to predict a class label from a collection of training data [9]. For a classification algorithm, one of the most important aims is to maximise predicted accuracy in scenarios that were not encountered during training.

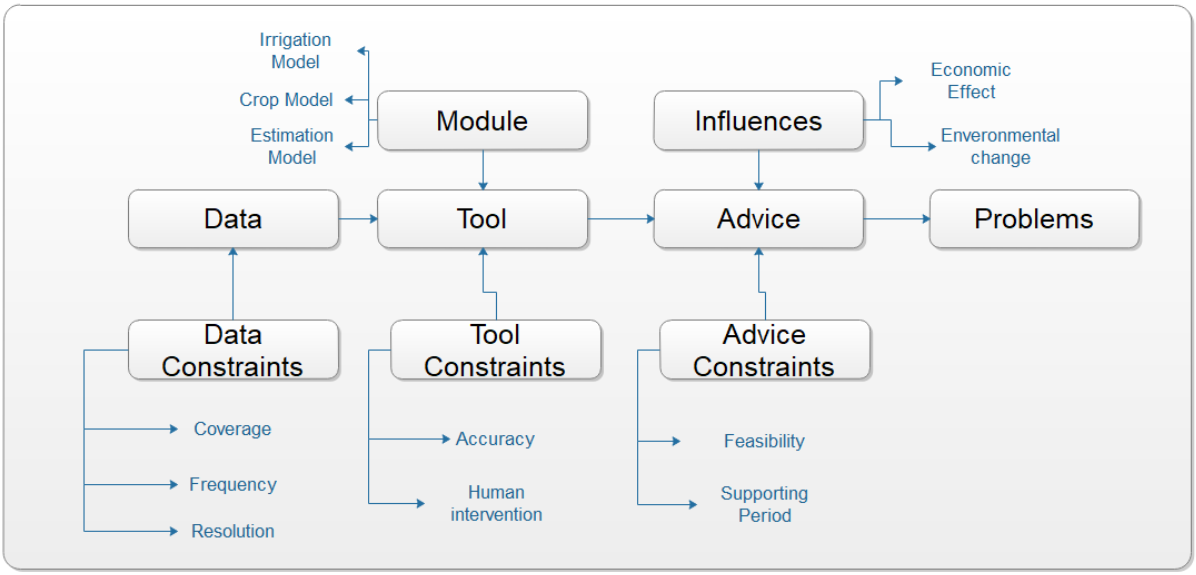

Figure 1: Data Analysis classification block diagram

Agriculture's raw data is extremely diversified. For the establishment of an agricultural information system, it is important to gather and store data in an organised manner and integrate it. Because of this, information and communication technology (ICT) must be used to extract important data from agriculture in order to discover trends and learn more about the subject.

Table 1: Parameter Description in Agriculture dataset

Dimensions	Description
Location	Contains description about the state, district, block, village and longitude and latitude positions
Crop	Name of the crop, its type and properties of crop
Farmer	Consists of information about the formers
Market	Contains description about location of the markets, price details of each crop at different time period
Soil	Types of soil, physical, chemical and biochemical properties of soil, location of soil
Water	Properties of water at different locations
Pesticide	Contains information about pesticides, it does level for different crops for different insects
Time	Year, quarter, month, day_of_week and time_key

This strategy will also lessen the amount of manual labour required. Lowering the cost of production, increasing output, and raising the market price are all possible outcomes of extracting data from electronic sources and transferring it to a secure electronic documentation system [10]. Data mining techniques can be used to gather crop information, allowing agricultural enterprises to foresee changes in customer conditions or behaviour. Finding links between seemingly unrelated bits of data requires a multi-faceted approach to data analysis. Farm data's computational demands and data mining's potential as a tool for knowledge management should be considered by researchers. Databases can store agricultural data, making it easier to manage transactions, retrieve information, and do analyses on that data [11].

3. Steps to effective decision making in agriculture

A decision-making process entails identifying a problem, gathering data, and considering the advantages and disadvantages of several solutions [12-16]. In order to make better decisions, it's important to organise relevant information and identify options in a step-by-step decision-making process.

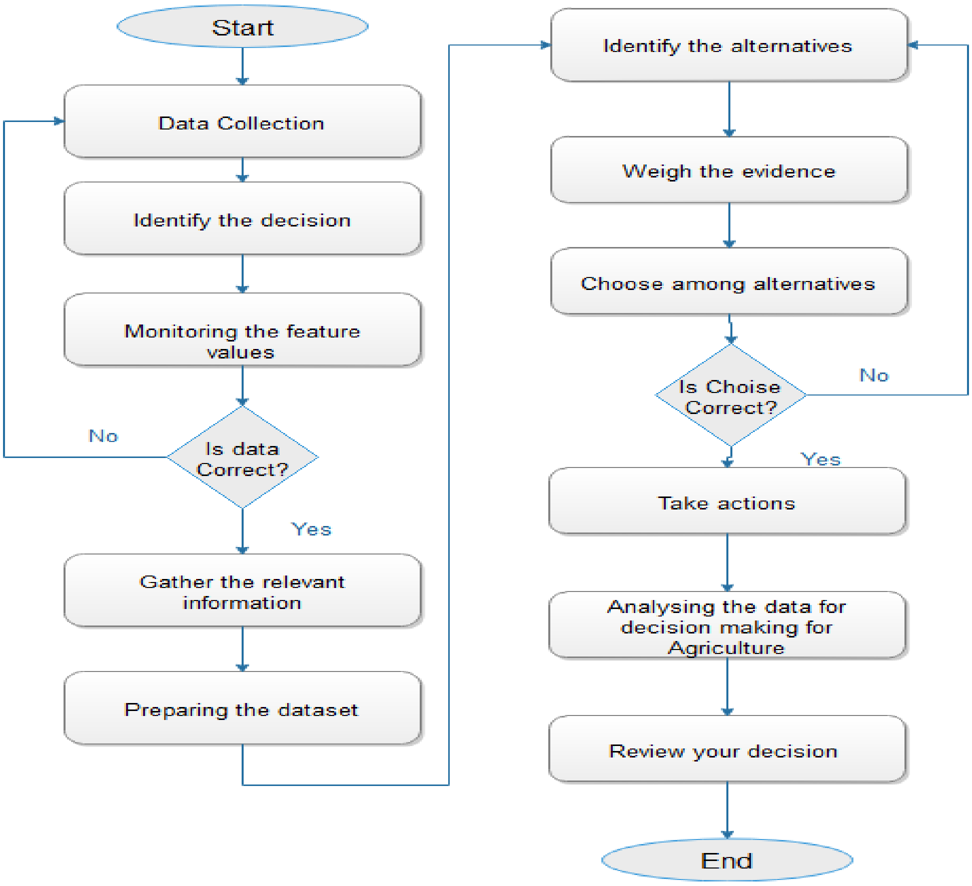

Figure 2: Flowchart for effective decision making in agriculture

Using this method increases the likelihood that you'll choose the finest alternative that is now accessible.

- Data Collection
- Identify the decision
- Monitoring the feature values
- Gather the relevant information
- Preparing the dataset
- Identify the alternatives
- Weigh the evidence
- Choose among alternatives
- Take actions
- Analysing the data for decision making for Agriculture
- Review your decision

It is critical that agricultural decision support systems (AgriDSS) not only provide up-to-date and relevant information, but also do so in a manner that minimises the negative environmental impact of agriculture [17-21]. It is widely acknowledged that the DSSs currently available are not being used to their full potential by farmers, advisers, specialists, and policymakers alike. For example, they are unable to recognise and appreciate the practical decision-making of farmers, which is one of the factors contributing to their failure.

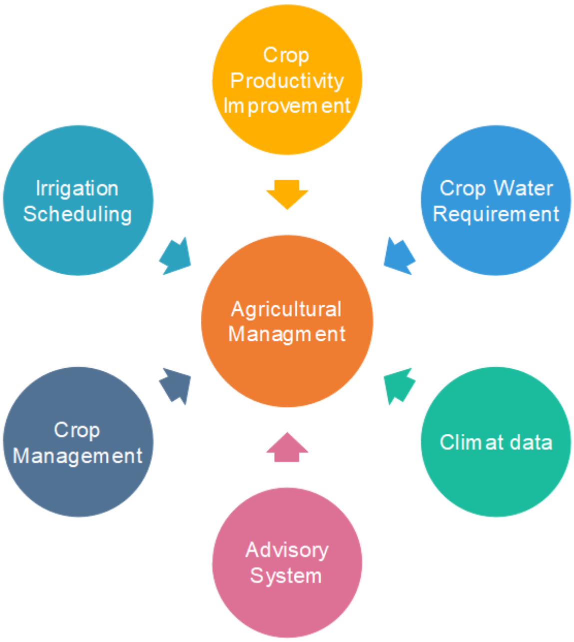

Figure 3: Activities involved in Agriculture Management

They are not equipped to deal with the high amount of complexity that comes with making judgments about long-term land use and development planning. Farmers are having a difficult time embracing these systems since they are based on what scientists and system developers believe is necessary in the current environment [22]. As a result, it is necessary to establish new relationships and foster greater understanding among agricultural stakeholders. User-centred design (UCD) has the potential to address the primary difficulties facing most DSSs since it places the farmers' experience at the centre of attention and engages them early and regularly in the design process, as opposed to traditional approaches.

Agriculture production and processing systems have gotten increasingly complicated as a result of the participation of biological, chemical, and physical processes such as soil, water, meteorological scenarios, and crop management strategies, to name a few examples of these processes. When used properly, a Decision Support System (DSS) provides a structured framework within which complicated systems can be represented, allowing them to be more easily comprehended while also allowing for the extraction of more information and fresh insights. It is an interactive computer-based expert system that assists decision makers in utilising data and models to address unstructured problems by guiding them through the decision-making process [24]. It is possible to contribute to the long-term sustainability of agricultural resources through the appropriate application of successful decision support. The activities in agriculture management can be grouped into many categories based on the important characteristics in agriculture such as the type of soil, the type of seed, the kind of irrigation, the type of fertilisers, and the type of meteorological data (Figure 3). Each of these operations requires the use of a decision support system in order to ensure successful and sustainable agriculture management.

4. Methodology

In the first place, agricultural data should be collected and processed as inputs to decision-making instruments rather than as end products (modules). The findings of computer simulations are used to produce agricultural management recommendations. The farmers can then choose the most appropriate option and put it into action [25]. Constrained resources must be taken into consideration in order to ensure that clients receive high-quality advice.

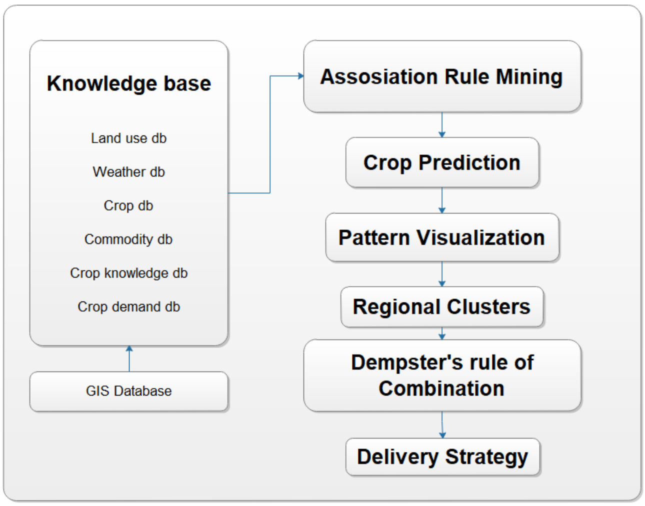

Figure 4: Block diagram decision making Module

First and foremost, the data in the knowledge base has been pre-processed to reflect logistical movement. So as a result, an association rule mining component collects relevant information from the data in order to identify interesting association rules [26]. The Apriori approach is used in the association rule mining component to find possible connections and assign a weight to each one in the association rule mining component.

Logistics managers can specify product types, quantities, and modes of transportation in a new delivery request that has been submitted. After receiving a new case, Dempster's rule of combination can aggregate connected relationships between instances and design the most efficient logistical route depending on the weights provided to each instance [28]. Dempster's rule of combination is a rule of thumb for combining two or more things. According to the QSDSS output, the route with the highest weight is the most optimal delivery strategy. Finding new, interesting, and useful data in a non-trivial manner is not an easy undertaking. It takes time and effort. In order to develop control mechanisms for exploration and rationalisation, data stored in various types of repositories, such as files and databases, is becoming increasingly important.

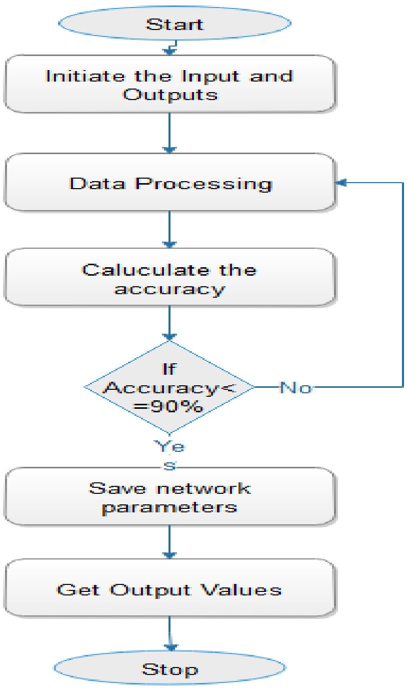

Figure 5: Flowchart for Accuracy Detection

Data is becoming increasingly necessary in order to create control mechanisms for exploration and rationalisation. In addition, useful information can be obtained from the data, which can be used to assist in decision-making processes. When viewed as a step in an iterative process of uncovering new knowledge, data mining is illustrated in the illustration in figure 5.

Crop demand analysis using Decision tree approach

The Decision Tree and the Weighted Decision Tree are two of the most important decision-making approaches used in crop yield analysis. The empirical equation for crop yield coefficients and the cost benefit ratio was derived using an analytical breakpoint (m) based on the optimization of a mean of three years of Tomato, Brinjal, Potato, and Bendi crop yields, as well as the cost benefit ratio.

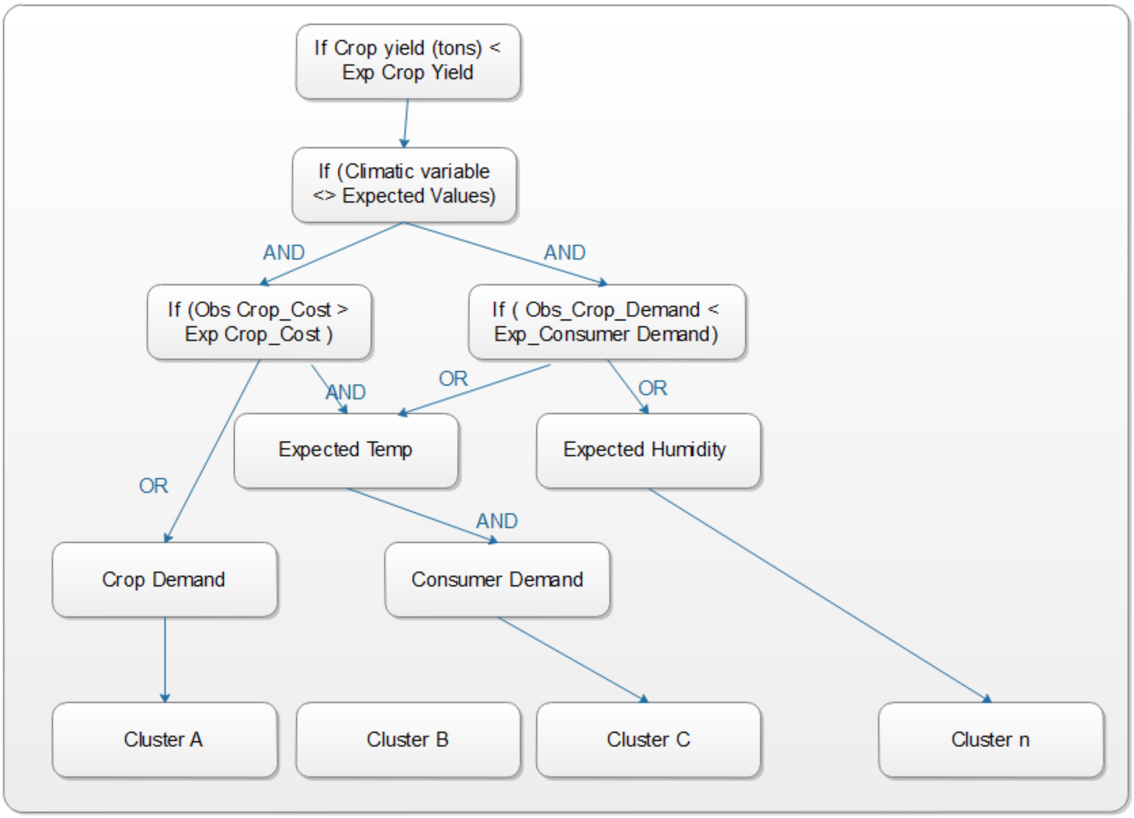

Figure 6: Decision tree approach for Crop demand Analysis

Methodology uses polygon analysis to mine related datasets, which consists of 3 steps:

- To generate and support on multi-dimensional clusters for variable crops, cost, region related datasets
- To adopt Meta clusters for inter co-ordinate of crop as components and parameters as weighted edges.
- To analyze on interesting patterns of crop growth using association rules
- To create summaries from clusters for crop yield and commodity sales

Crop yield Decision making analysis (CDMA)

This is illustrated in Fig. 7, which collects agricultural growth parameters from the data set and calculates the commodities cost-benefit ratio using a CDMA functional architecture. The baseline dataset has been confirmed and established on the basis of historical data that has already been collected and validated against standard datasets. A set of metrics (Table-1) based on crop growth characteristics and commodity sales are utilised to create the cluster structure. The CDMA supports the invariance of agricultural productivity, growth, and the cost-benefit features of business invariance that are variable.

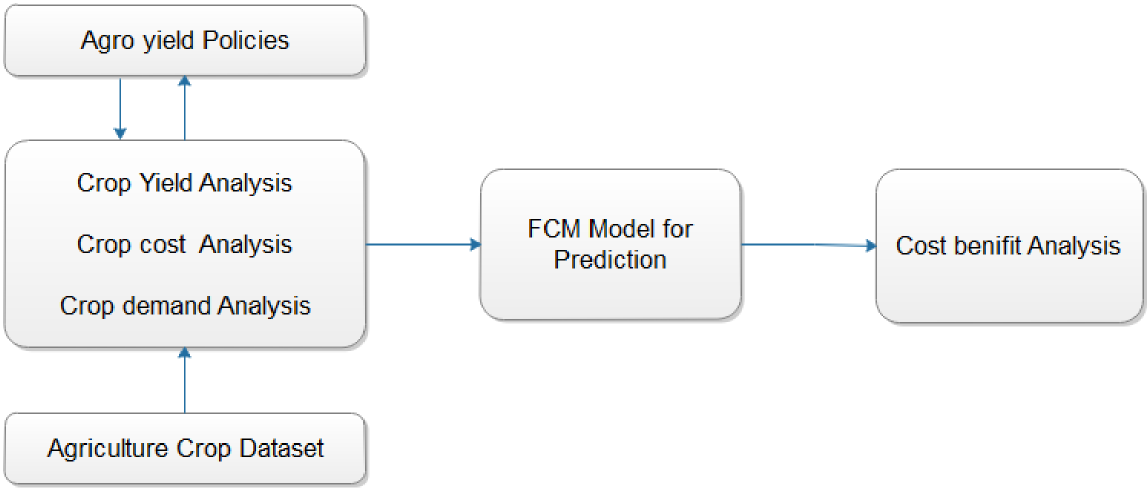

Figure 7: Block diagram for cost benefit analysis using FCM Model

5. Analysis and Results

When evaluating the outcomes, the predicting MAPE metrics were utilised, as illustrated by Table 2. A statistical tool, WESSA, was utilised to conduct MLR experiments in the same manner as before. Using the WESSA FCM Tool, it was essential to apply pre-processed data in order to derive regression equation estimates, which were then assessed by varying the parameters of a single regression equation. When the forecast price was the dependent variable in this equation, there were a slew of meteorological parameters, such as temperature, precipitation, and humidity, that were the independent variables.

Table 2: Agriculture Commodity of Bendi, Tomato, Brinjal and Potato in different seasons

		Winter		Summer		Monsoon	
		WESSA	NN	WESSA	NN	WESSA	NN
Potato	Training	6.09	5.23	9.02	8.54	10.32	9.54
	Testing	10.11	15.09	14.62	17.82	13.02	16.94
	Growth	1.54	1.23	1.32	1.67	1.54	1.32
Brinjal	Training	5.23	6.09	10.32	9.54	9.02	8.54
	Testing	14.62	17.82	15.09	13.02	16.94	10.11
	Growth	1.23	1.54	1.67	1.54	1.32	1.73
Tomato	Training	9.02	8.54	6.09	5.23	9.54	10.32
	Testing	16.94	13.02	10.11	17.82	15.09	14.62
	Growth	1.54	1.67	1.32	1.73	1.23	1.82
Bendi	Training	5.23	10.32	9.54	6.09	8.54	9.02
	Testing	15.09	14.62	17.82	10.11	16.94	13.02
	Growth	1.73	1.23	1.67	1.32	1.54	1.45

Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 4 Issue 2 (2024)

Figure 8: Graph for Agriculture Commodity of Bendi, Tomato, Brinjal and Potato in different seasons

Each season was further divided into four months to improve the accuracy level. Every season's model was tested, and the results showed that the monthly model provided accurate results. Additionally, the price of agricultural products is influenced by seasonal and other factors in the agricultural commodities market. A year's worth of data may be seen in the accompanying Table 3, which breaks the year down into three distinct seasons: winter, summer, and monsoon, with one month of data for each season. Seasonal models (one model per season) were used to conduct experiments on agricultural product Bhindi (Ladies Finger). The results are shown in Table 3. Here's how precise we were able to get:

Table 3: Seasonal wise Accuracy for crop

Season	Data Analysis Months	Accuracy (%)			
		Potato	Brinjal	Tomato	Bendi
Monsoon	July to Oct	87.43	82.92	85.32	86.56
Summer	Mar to June	75.32	75.34	78.65	71.65
Winter	Nov to Feb	93.25	91.67	93.45	88.27

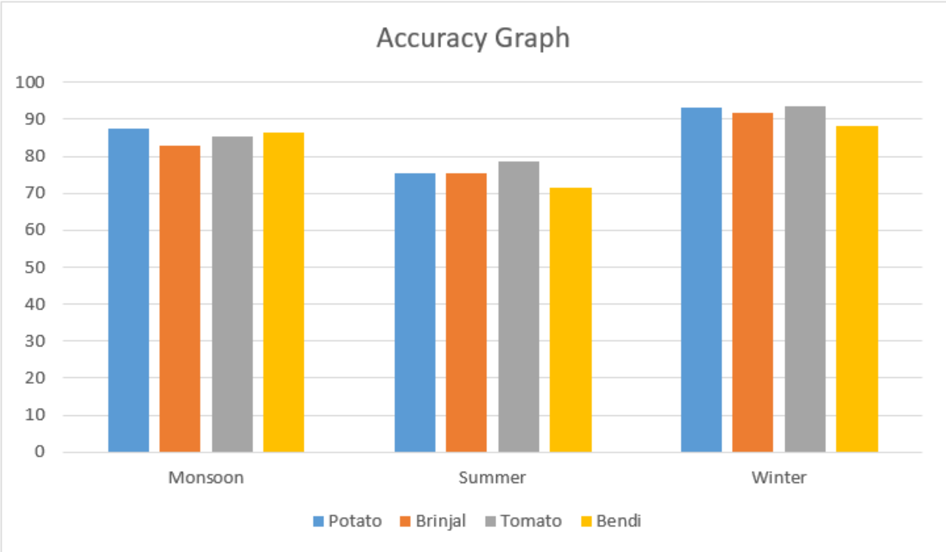

Figure 9: Seasonal wise Accuracy graph for Potato, Brinjal, Tomato and Bhindi

6. Conclusion

Several factors influence a company's choice to enter the market. The driving force behind crop sales and, in a sense, the commercialization of agriculture is discovered through a two-step approach including correlation and regression. Correlation tests were conducted to examine the association between crop sales, seed source, fertiliser use and crop area, owned and leased land, loan, household size, education and total livestock, and the age of the farmer. The selling of the entire crop was shown to be associated with the first four variables. While the sale of two crops was shown to be marginally linked with land ownership and leasehold interest. To improve market price forecasting accuracy after opting on a monthly method, research were done. The implementation of this mathematical model has been presented in the form of a comprehensive graphical user interface (GUI) for the benefit of the reader. The results of the yearly model are presented first, and then the seasonal and monthly approaches are discussed. According to this finding, the Monthly technique is more accurate than the other two approaches, hence it is the best of the three. The comparative examination of monthly approaches with and without the fitness factor is find out in further work.

References

[1] H. Wang, "Empowerment of Digital Technology to Improve the Level of Agricultural Economic Development based on Data Mining," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1111-1114, doi: 10.1109/ICICCS51141.2021.9432369.

[2] S. A. Lokhande, "Effective use of Big Data in Precision Agriculture," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 312-316, doi: 10.1109/ESCI50559.2021.9396813.

[3] H. Gao, "Agricultural Soil Data Analysis Using Spatial Clustering Data Mining Techniques," 2021 IEEE 13th International Conference on Computer Research and Development (ICCRD), 2021, pp. 83-90, doi: 10.1109/ICCRD51685.2021.9386553.

[4] R. Gupta et al., "WB-CPI: Weather Based Crop Prediction in India Using Big Data Analytics," in IEEE Access, vol. 9, pp. 137869-137885, 2021, doi: 10.1109/ACCESS.2021.3117247.

[5] G. Weikmann, C. Paris and L. Bruzzone, "TimeSen2Crop: A Million Labeled Samples Dataset of Sentinel 2 Image Time Series for Crop-Type Classification," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 4699-4708, 2021, doi: 10.1109/JSTARS.2021.3073965.

[6] N. Chergui, M. Kechadi and M. McDonnell, "The Impact of Data Analytics in Digital Agriculture: A Review," 2020 International Multi-Conference on: "Organization of Knowledge and Advanced Technologies" (OCTA), 2020, pp. 1-13, doi: 10.1109/OCTA49274.2020.9151851.

[7] C. N. Vanitha, N. Archana and R. Sowmiya, "Agriculture Analysis Using Data Mining And Machine Learning Techniques," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019, pp. 984-990, doi: 10.1109/ICACCS.2019.8728382.

[8] K. Soma, M. Bogaardt, K. Poppe, S. Wolfert, G. Beers, D. Urdu, M. P. Kirova, C. Thurston, and C. M. Belles, "Research for agri committee - impacts of the digital economy on the food chain and the cap. policy department for structural and cohesion policies," European Parliament. Brussels, Tech. Rep., 2019.

[9] H. V. Nguyen, M. A. Naeem, N. Wichitaksorn and R. Pears, A Smart System for Short-Term Price Prediction using Time Series Models, Computers and Electrical Engineering, 76 (2019), 339–352.

[10] A. Vohra, N. Pandey and S. K. Khatri, Decision Making Support System for Prediction of Prices in Agricultural Commodity, Proceedings of the IEEE Amity International Conference on Artificial Intelligence (AICAI), (2019), 345-348.

[11] J. Gholap, A. Lngole, J. Gohil, Shailesh and V. Attar, Soil Data Analysis using Classifcation Techniques and Soil Attribute Prediction, International Journal of Computer Science, 9 (3) (2018), 1–4.

[12] V. Patodkar, S. Sharma, S. Simant, C. Shah and S. Godse, E-Agro Android Application (Integrated Farming Management Systems for Sustainable Development of Farmers), International Journal of Engineering Research and General Science, 3 (1) (2018), 368-372.

[13] H. Anandakumar and K. Umamaheswari, "A bio-inspired swarm intelligence technique for social aware cognitive radio handovers," Computers & Electrical Engineering, vol. 71, pp. 925–937, Oct. 2018.

[14] D. Elavarasan, D. Vincent, V. Sharma, A. Zomaya, and K. Srinivasan, "Forecasting yield by integrating agrarian factors and machine learning models: A survey," Computers and Electronics in Agriculture, vol. 155, pp. 257–282, 2018.

[15] D. Patricio and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review," Computers and Electronics in Agriculture, vol. 153, pp. 69–81, 2018.

[16] S. Sabzi and Y. Abbaspour-Gilandeh, "Using video processing to classify potato plant and three types of weed using hybrid of artificial neural network and particle swarm algorithm," Measurement, vol. 126, pp. 22– 36, 2018.

[17] I. Oliveira, R. Cunha, B. Silva, and M. Netto, "A scalable machine learning system for pre-season agriculture yield forecast," in the 14th IEEE eScience Conference, 2018.

[18] L. Kouadio, R. Deo, V. Byrareddy, J. Adamowski, S. Mushtaq, and V. P. Nguyen, "Artificial intelligence approach for the prediction of robusta coffee yield using soil fertility properties," Computers and Electronics in Agriculture, vol. 155, pp. 324–338, 2018.

[19] A. Wang, C. Tran, N. Desai, D. Lobell, and S. Ermon, "Deep transfer learning for crop yield prediction with remote sensing data," in Proceedings of the COMPASS'18, Proceedings of the 1st ACM SIGCAS conference on Computing and Sustainable Societies. Menlo Park and San Jose, CA, USA. June 20-22, 2018.

[20] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep gaussian process for crop yield prediction based on remote sensing data," in the Thirty First AAAI Conference on Artificial Intelligence. AAAI Publications, 2017, pp. 4559–4566.

[21] P. Filippi, E. Jones, T. Bishop, N. Acharige, S. Dewage, L. Johnson, S. Ugbaje, T. Jephcott, S. Paterson, and B. Whelan, "A big data approach to predicting crop yield," in Proceedings of the 7th Asian-Australasian Conference on Precision Agriculture 16–18 October 2017, Hamilton, New Zealand., 2017.

[22] A. Kamilaris, A. Kartakoullis, and F. Prenafeta-Boldu, "A review on the practice of big data analysis in agriculture," Computers and Electronics in Agriculture, vol. 143, pp. 23–37, 2017.

[23] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," Geoscience and Remote Sensing Letters, vol. 14, no. 5, pp. 778–782, 2017.

[24] M. Kaur, H. Gulati and H. Kundra, Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications, International Journal of Computer Applications, 99 (12) (2017), 975 – 8887.

[25] R. Shirsath, N. Khadke, D. More, P. Patil and H. Patil, "Agriculture decision support system using data mining," 2017 International Conference on Intelligent Computing and Control (I2C2), 2017, pp. 1-5, doi: 10.1109/I2C2.2017.8321888.

[26] Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data", Journal of Big Data 2017, DOI 10.1186/s40537-017-0077-4, 5 July 2017

[27] Priya Nagose1, Ankita Belkhode, "Efficient Data Mining for Increasing Agriculture Productivity", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 5 Issue III, ISSN: 2321-9653, page 281-282, March 2017.

[28] Majumdar J, Ankalaki S. "Comparison of clustering algorithms using quality metrics with invariant features extracted from plant leaves", Advanced Science Letters, Volume 23, Number 11, pp. 11211-11216(6), November 2017.