DOI: http://dx.doi.org/10.26483/ijarcs.v17i2.7414

Volume 17, No. 2, March-April 2026

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

USE OF DATA MINING TECHNIQUES IN AGRICULTURE

Chandra Shekhar Patel
Sardar Patel University
Balaghat, (M.P.)

Krishna Kumar Verma
Government College Tala,
A.P.S. University, Rewa (M.P.)

Tryambak Hiwarkar
Sardar Patel University
Balaghat, (M.P.)

Abstract: In modern era, agriculture sector has many challenges facing by farming communities. Numbers of services with ICT tools are being provided by government/ private leading departments to support agriculture activities and increase productivity but still needs to indentify information gaps that may help in effective decision making. There are need to address the efficient use of data mining techniques so that effective and accurate decision making can be done. This study presents a review of literature done by different authors. Experiments related to supervised learning (classification) techniques has also been done. Content which has been covered in literature review are application of data mining, machine learning, artificial intelligence in different sub fields of agriculture like disease detection, yield prediction, crop quality etc. Classification algorithms have been used to classify different types of crops. Five classification algorithms NB, SVM, NN and Decision Tree have been selected. Comparison of different parameters such as Accuracy, Kappa Statistic, RMSE and so on related to classification model have been done. It has been found that the decision tree (PART) algorithm is more suitable than other classifier on selected dataset.

Keywords: Data Mining, e-agriculture, data analysis, classification algorithm

I. INTRODUCTION

In agriculture, modern farming has been mapping number of hardware as well as software technologies used in Agri Production. It has incorporated ICT, Cloud Computing, and Mobile Computing etc. with additional backbone of DM/ML/AI analysis techniques. However, it still faces several challenges such as massive increase in food demand due to the global burst of human population along with climate change [1], depletion of natural resources by agro ecosystems [2], the dilemma between food choices and environmental concerns [3] as well as some health and safety related issues[4]. It also goes through some other dataset problems like storing, handling, managing, analyzing a large scale multidimensional / scattered dataset. Reducing the environmental burden by increasing agricultural output per land area and adopting greener technologies that require less resource is necessary to counteract some of these issues burdening agriculture. All this is possible with the help of basic methodologies which support in evolution of agriculture to precision agriculture. Future sustainable agriculture will enable maximum production with safety of the environment [5].

Overall, the technology involves four pillars of smart agriculture to cater to the increasing requirements (a) natural resources management, (b) ecosystem conservation, (c) service provisioning and deployment techniques and (d) modern-day technologies utilization[6]. Big Data generated from digital technologies, demanding high storage capacity and editing, analytics and interpretation. This will potentially have a high social and environmental value, but also for the decision makers [7]. But the reality is that big data has 5-V requirements as they are named; quantity, variety, speed and accuracy and value [8]. The traditional / Old data analysis methods are not enough to meet the evolving needs of today's smart agriculture leading to hampering the essentials and insights out of big huge real time data [9]. To tackle this, Machine Learning (ML) a subset of Artificial Intelligence (AI), has been developed leveraging the existence potential for exponentially more powerful computing power today than in the past[10]. Agriculture in word level Data Mining/AI applications, analysis of the literature [11], concerning 2004–2018, has listed four basic types. The categories were crop, water, soil and livestock management. In detail, crop Management articles were the majority in all categories 61% of total articles then classified into:

- Disease detection;
- Yield Prediction;
- Weed Detection;
- Crop Recognition;
- Crop Quality.

Indian farming is based on the weather, weather changes, insect attacks and production from year to year. Due to which output of farming in rural India is unstable as farmer do not understand what will grow effectively in the next growing/ upcoming season. Hence, farmers and other related decisions makers not ready for weather condition will make a decision base with the local/ geographical weather situations as they know. Climate risk, with pest variations on top, mostly brings forward old-style big picture approaches resulting in compensating for some lost farming effort to prevent sour years from turning out worse. Farming practices can be rearranged to reduce unproductive cost and capitalize on favorable forecast conditions if improved predictions of weather-related parameters or crop yields are given four to five months prior. Genetically human population is increasing with it and hence there is need to provide food/ grain for Human being. So, the use of data is also happening in farm sector which ultimately give increase to net productivity by using various technologies and their hardware & software. Moreover, terrible consequences of climate changes have long since made the advancement toward modern farming by means

© 2025-2027, IJARCS All Rights Reserved

of leading technologies a necessity. Therefore, the agricultural and livestock sector in this period started to adopt more advanced technologies such as machine learning, information, and communication technologies agents to add production. Thus, the idea of precision agriculture was introduced that revealed a new face of agricultural science. Otherwise, agriculture is still largely an indirect source of humanity, and it provides the basis for contemporary farming employment by work force in most countries. This will help to get smart agriculture practice as India is the most populous countries in the world that have many of people depends upon to farming. This means that in India 18% of GDP, and half of the population capable can work through these companies. Dhan, Saraso& Gehu are the main indigenous crops of India have a total net sown area of 70% in the country (Source: "Ministry of Statistics and Program Implementation").

This study deals with review of literatures done by different authors. Content which has been explained in this paper has included the applications of DM/ ML/ AI in field of agriculture sector. Additionally, classification techniques have been utilized to classify "type of crops" dataset.

II. RELATED WORK

The concept of e-agricultural which makes the use of ICT tools and other technology to increase the productivity in agricultural sector has been discussed. ICT helps to store agricultural related information that can be utilized for further data analysis in research. Authors have reviewed literatures by covering the benefits of Cloud Computing as ICT technology. The study has been done for Kenyan agriculture. It has been concluded that mobile based services for agricultural sector is very useful for taking efficient decision which help to increase agricultural productivity. Radio and TV are still commonly used by rural people. Also smallholder farmers need to be trained and empowered so that they able to utilize new ICT/ mobile enabled agricultural technologies [12].

A study of *Nigerian agricultural ecosystem* has been presented by Umar, et al. (2022). They have carried out systematic review that involves the study related to applications of AI, ML and DL in the field of agricultural sector. Review has been classified in different sub field of agriculture that has adopted ML/DL technology. These categories were Crop Management, Livestock Management, Water Management and Soil Management. Analysis showed that 35% researcher have used DL whereas 65% researcher have used ML technology to perform data analysis tasks. ANN was mostly used algorithm among others. Authors have also discussed opportunities, Potentials and challenges in this area. Major challenges in this field are lack of enough and quality full dataset, Poor Infrastructure, Privacy and Security Issues etc [13].

Modernization in agriculture sector can be possible by using new technologies. In China's agriculture, this concept has been covered by transforming agriculture practice into inorganic agriculture. To cover this concept Wu & Ma (2022) have discussed that China's agriculture has transformed into inorganic agriculture. Image target detection methods have been used in pest target detection and recognition. Deep learning along with the perspective of block chain has been proposed to analyze agricultural data. Through comparative study authors have observed that average precision of GA-CPN-LAR increases by 4.2% and under ResNet-50 networks, AP of this algorithm was significantly better than other methods. Recall and Precision comparison of SSD, GA_CPN-LAR, FPN and Faster R-CNN methods over MPD dataset have been presented. Implementation architecture of proposed method is also discussed in research. This has included data layer, NN detection layer, Consensus layer and application layer. Authors have concluded that proposed method have great application potential in agriculture sector [14].

The use of machinery in farming or cultivation can reduce labour and debt. Classification of effective use of machines in farming has been done using ML algorithm. On the basis of use of farming equipment; farmers have been divided into three groups: small, moderate, and large. K-nearest neighbors (KNN), logistic regression (LR), and decision trees (DT) have been used in classification task. The performance of DT and KNN were better than other method but researcher have selected DT because it is a non distance based algorithm. The study has been done on data collected from 562 farmers in Punjab through open-end questionnaire. Data collection has been done over 12 different parameters in which classification has been done on "Aware of new farming tools". It has been calculated that 67.06% farmers was under small category, 31.08% farmers was under moderate category and 1.06% farmers was under large category [15].

Effective use of water in irrigation is very necessary which can be achieved by ML and Internet of Thinks (IoT). IoT driven model for irrigation has been proposed by Singh et al., (2022). Most important technologies in proposed system were IoT and long-range, low power WSN, and ML algorithm that are used in prediction of irrigation scheduling. The selected dataset was of 500 observations. The parameters of soil and weather dataset was "AirTemperature", "Humidity", "WindSpeed", "Volumetric and Moisture Content (VMC)", and "SoilTemperature". Class level has been developed which contains 0 and 1 and included as sixth column in dataset. Value 0 and 1 represent whether irrigation is necessary or not, depending on soil and weather conditions. Six different ML algorithms has been selected for doing experiment which are: linear logistic regression (LR), LDA, non liner KNN, classification and regression tree (CART), Gaussian naïve Bayes (NB) and support vector machine (SVM). It has been observed that SVM, KNN and NB algorithms are frequently employed in agriculture applications. Also LDA algorithm has given highest accuracy of 91.25% amongst all six ML algorithms. The accuracy of other algorithm was LR- 82.5%, KNN- 83.75%, CART 85.0%, NB- 85.0% and SVM 63.75%. It has been concluded that proposed intelligent irrigation DSS can be utilized by farmer for effective decision making in their farming practices [16].

The use of ICT, AI tools and new techniques in support of improving agricultural productivity has been discussed by Songol et al., (2021). Authors have argued that leveraging digital tools can help to farmers for improving the farms production output. Decision making process could be improved with the help of emerging technologies like AI and ML. These techniques may provide timely and accurate agriculture related information that can be utilized in precision agriculture. It was also discussed that AI may play a crucial role in ensuring pests control, soil selection and reducing knowledge gap between the technology and farmers [17].

In an article [18] application of AI for the food industry has been explained. Data science with AI can improve the quality of services such as online delivery, food outlets, cafes,

© 2025-2027, IJARCS All Rights Reserved

and so on. It has been discussed that by use of AI and ML based system, food production and delivery processes can be efficiently handled. "Application of ML in restaurant business", "AI in Food safety", "Remuneration of AI in the food industry", and "Future application of AI in Food Industry" was the major discussion of this article.

The big challenge for the agri food industry is the population of world which is expected to reach larger than 9 billion by 2050. Along with the challenges of agricultural requirement; authors have reported the importance of AI and ML as a predictive approach to improve productivity in the field of food and agriculture. Limitations of these techniques have also been discussed that should be considered by data analyst. Four stages of agriculture supply chain have been discussed that are "preproduction", "production", "processing" and "distribution". It has been discussed that for good production in near future it is necessary to use or adopt digital technologies such as sensors, machinery, monitoring of crops etc [19].

Precision Agriculture (PA) is the adoption of ICT for farm management to improve productivity by using the effective practices in agriculture. The causes that affect agriculture productivity are soil degradation, increasing costs and weather variation. To overcome these challenges PA uses sensor technology. Also ML and AI have game changed PA by multidirectional way. Comprehensive survey has been done by author Singh et al. (2021) and proposed a multidisciplinary "*AgriFusion*" IoT based architectural solution. Proposed architecture will provide facility for integration of multi-disciplinary technologies related to agriculture. Farm management industrial solution and their important technology have been discussed in detail. Additionally authors have proposed the solution architecture for designing the Key Performance Indicators (KPI) [20].

Integration of various distributed satellite and climate data for developing prediction model to predict wheat yield has been done by Cai et al., (2019). Australian geographic wheat data from 2000 to 2014 has been selected to do this work. Researcher has adopted regression method LASSO and three ML methods (SVM, RF, and NN) to build yield predictive model for wheat crop. Crop productivity has been approximately measured by use of EVI "enhance vegetation index" from MODIS, SIF "solar-induced chlorophyll fluorescence" from GOME-2 and SCIAMACHY. Authors have found that non-linear ML methods i.e. SVM, RF and NN have performed better than linear method. The comparison of SIF and EVI satellite data showed that better performance can be achieved by using EVI at SD scale. The study had some limitations one of them was use of static wheat map for all considered year which change from year to year. Authors have observed that by use of proposed method, we can achieve timely and optimal solution for wheat yield-prediction. In future the proposed method will be used for different regions for different crops [21].

Durga et al. (2019) used HOG operation and done classification task. In this work tomato & maize Indian crops have been selected to detect plant diseases. Researcher used HOG operation and performed classification task. Plant diseases have been classified using SVM and ANN classifier. Dataset used has 200 leaf images of tomato and maize. Sixty four input layers, 64 hidden layers and 03 output layers have been used in ANN. The Python API -keras has been used in this experiment. Classification accuracy for tomato crop using SVM was60-70% and using ANN accuracy was 80-85%. Also the classification accuracy for corn crop using SVM was 70-75% and using ANN was 55.65. It has been concluded that accuracy may be vary for different types of databases [22].

There are many factors that may affect the productivity of crops but the most important factor is plant disease which can be identified using leaf image analysis. Authors have discussed the use of computer vision and ML techniques to achieve their goal. KNN classifier with GLCM texture features has been used to classify plant leaf disease images. The proposed technique has successfully detected and recognized plant diseases with 96.76% accuracy. Developed model able to classify different types of diseases like "alternaria", "alternata", "anracnose", "bacterial" and so on. The experiment has been simulated on MATLAB. The proposed model has also been compared with existing model. In future, classification will be done for more diseases. Also research will deploy NN algorithm for improving the accuracy of classifier [23].

Crop yield prediction is a new method which may help in early farming activities. Maimaitijiang et al. (2019) have developed a new unmanned aerial vehicle (UAV) platform that helped in efficient data collection at very high and spectral resolutions. The objective of this study was to evaluate the proposed UAV system using RGB for soyben grain. DNN algorithm has been used to analyze dataset. Multimodal information has been extracted and used to predict crop. PLSR, RFR, SVR, DNN-F1, and DNN-F2 were used in prediction task. It has been found that multimodal data fusion technique has improved the accuracy of yield prediction. Also DNN-F2 has given highest accuracy with R2 of 0.720 and RMSE of 15.9%. However, there is need to test efficiency of this approach for other crop types at different geographic and environmental situations [24].

Soil moisture is directly affected by environmental parameters such as air humidity, air temperature, UV, and so on. An IoT based smart irrigation architecture that has used a hybrid machine learning algorithm to predict the soil moisture has been proposed [25]. The proposed architecture has seven main components. These components have been grouped into three layers which are data collection & transmission, data processing & intelligence and application layer. Past sensor's data and weather forecasted data have been used in prediction task. Support Vector Regression (SVR) and k-means clustering ML algorithm has been combined to observe the changes in soil moisture. Proposed hybrid algorithm has given good accuracy with minimum Mean Squared Error. Correlation coefficient (R) value of predicted Soil Moisture Differences (SMD) using old SVR was 0.3134 whereas the value of R given by proposed algorithm was 0.5592. Result of predicted soil moisture has been used to develop the smart irrigation scheduling algorithm. Interactive web portal has also been developed to visualize results. Future work is to conduct a water saving analysis using proposed algorithm which will also include the minimization of the system cost.

Development of CNN model to detect & diagnose plant disease has been done. In this experiment, healthy and diseased plant leaves images have been used. Open dataset of 87,848 images have been selected that has 25 different plants and 58 different classes of plant diseases considering healthy plants. The learning rate of model was started from 0.001 & descending every 20-epochs by1/2 or 1/5. Accuracy of proposed model was 99.53% [26].

© 2025-2027, IJARCS All Rights Reserved

Random forest technique has been used to classify diseased and non diseased leaf images. Features from images have been extracted using "Histogram of Oriented Gradient (HOG)". Logistic regression, SVM, KNN, CART, Random Forest and Naïve Bayes ML models were developed and compared. Among all models Random Forest model was found best with 70.14% accuracy [27].

Dhakal et al. (2018) performed leaves image classification to detect diseased infected leaves. In this process, dataset was collected and initially processed using image processing techniques. Four labels were used which were Bacterial_Spot, Yellow_leaf, Curl_Virus, Late_Blight and Healthy_Leaf. Neural Network with 20 epochs has been used. Among several ANN the best performance was 98.59% found. The performed test was efficient to detect plants diseases like bacteria, cural virus, & late blight [28].

Sindhu et al. (2017) Framework for ICT based agricultural system has been proposed and it has been discussed that in near future ICT based agriculture will be the need of farmers. ICT may provide facility in development of more sophisticated techniques for agriculture related problems that will lead in increase food production. In this study concept of image processing techniques, SVM, Bayesian Network, ANN and Association rule have been discussed under methods & techniques. Solis classification, soil fertility, use of GPS & GIS, optimization of pesticide control of weeds etc. have been discussed under application of ICT for e-agriculture [29].

Productivity of any crop can be increased by using smart agriculture technology. Kumar et al. (2017) proposed a precision system model for monitoring of sugarcane-crop. Three main parameters "temp.", "humidity" and "moisture" have been considered. KNN clustering method along with SVM classification method has been used to identify the infection through images of sugarcane crop. The accuracy of proposed model was 96% on selected 200 images. Testing was carried out at Lolai, Gomti Nagar Extension, Lucknow dataset. Details of temperature, humidity and moisture sensor have also discussed.The implementation of proposed system has been done using EAGLE7.1.O. Developed system has combined the process and data of sensors, image processing and classification algorithms. The developed system can be utilized in detection of crop diseases that will help farmers to take effective decision [30].

Nine machine learning algorithms KNN, SVM, NN, LR, Rule induction, Fast large margin, DT, RF, and LR have been used to predict soil moisture. Experiment has been performed on data set of Transylvania. The best prediction model was KNN with an accuracy of 74.36%. Test experiment has been performed on production data set using same algorithm. Expected moisture has been calculated with known historical data and predicted moisture has been calculated using ML algorithm. By comparing both value of moisture, error value has been calculated. The average error was -0.3% and absolute values was 0.68%. Authors have observed that proposed system may be used as warning system in agriculture practices. Similar research will be done for different locations using proposed process in future [31].

Rajeswari et al. (2017) discussed that use of Internet of Things (IoT) and mobile phone in field of farming activities has increased nowadays. Use of these technology generate large amount of data and to store these data there is need to use cloud storage mechanism. These data can be analyzed by using big data analytics. In this research, cloud based big data analysis has been proposed to analyze data. Expected outcome was identifying the fertilizer requirements, crop analysis, market & stock analysis. Authors have proposed a methodology which is a combination of IoT, cloud computing, big data analytics and pattern prediction. Predictive results are proposed to communicate the farmers in their mobile phone so that effective and efficient decisions can be made by farmers. The proposed model will also facilitate the estimation of total production per crop region. In future, the implementation of proposed method will be done [32].

Ramos et al. (2017) proposed a non-destructive method for counting the number of fruits on a coffee branch. In this work, authors have used digital images of a single side of the branch and paths of growing fruits. Machine Vision System (MVS) has been constructed, which able to count and identify harvestable & non-harvestable fruits from coffee branch images datasets, this system uses image processing algorithm to identify it. After this, linear estimation models were developed. Model was able to categorize fruits into three categories which are harvestable, not-harvestable and fruits. The propose system shown a correlation higher than 0.90 for yield prediction. The performed experiment has count error greater than 6%, which can be minimized by using appropriate image capturing [33].

Losses in crops can be reduced and best prices in market can be achieved using prediction techniques in agriculture sector. Sellam and Poovammal (2016) have developed a predictive model using regression analysis (RA) and environmental factors have been analyzed. The environmental parameters that have been considered were "Area under Cultivation (AUC), Annual Rainfall (AR) and Food Price Index (FPI)". These variable was used for analyzing the relationship & it's effect on yield in rice crop. Time period was 10 years from 1990-2000. Highest value for 'rsq' was found equal to 0.7242 for prediction of yield from AUC. It had been noticed that as soon as the cultivation area increases, the yield of crop also increases. The research will be extended by using other factors like MSP, CPI, WPI etc. in near future [34].

Awuor et al. (2016) have built an information access point as e-agricultural product for Kenya. The framework has been proposed to provide information related to agricultural practices to farmers in any time. It has been discussed that to develop such type of platform there is need to address some issues related to interest of managers, developers and users, and information related to sustainable agriculture. The proposed framework was four layers architecture that are "Foundation Layer, Management Layer, Data Layer and Information Access Layer". In last layer authors have also discussed the adoption of new technologies such as Machine Learning System, Big data etc. While developing this framework, two important issues has been concerned, first was the integration of local data center with private cloud in initial stage and second was high productivity and food security [35].

Identification of specific plant from 'leaf vein pattern' has been done using proposed deep learning (deep convolutional network) model. Four stage pipeline processes has been used in image dataset these are Vein segmentation, Central patch extraction, vein measures and classification. Under classification, three ML algorithms SVM, PDA, and RF have been tested. CNN configuration has been done using convolutional layer plus ReLU function plus dropout. The experiment run time was around two weeks for designing and running cross validation in which parameter optimization and

© 2025-2027, IJARCS All Rights Reserved

error estimation task has also been covered. It has been observed that accuracy has improved monotonically with the depth of the DL model. Hence it is concluded that depth may be a key element for solving this type of problem. Authors will explore other potential applications of DL like "weeds detection" or "seeds viability tests" [36].

Amatya et al. (2015) developed a machine vision system to segment and detect fully foliage branches of cherry tree. In this step, development of image segmentation has been done that cable to identify visible segments of the cheery tree branches. By use of Bayesian classifier, authors have classified images into four classes: branch, cherry, leafs & background. Bayesian classifier had 89.6% accuracy to identify branch pixels. After that result of previous step has been used to analyze and link individual sections of branch. To represent the particular branch linear and logarithmic model were fount as the best fit model. Overall the accuracy to detect branch in 141 test images was 89.2%. Experiment has been done using MATLAB. Derived results can be utilized to identify presence of full foliage & fruit in cheery tree during farming season. Identification of localized cherries, integration of 3D methods etc. shall be done in future [37].

E-agriculture is a concept that includes electronic techniques in agriculture practices and it is a platform which provides support for marketing of agriproducts with support of e-commerce. A study related to farmer's awareness in e-commerce has been conducted. It has been found that there is less level of awareness related to use of e-agriculture as e-commerce by farmers for marketing their products. Author has suggested that the cost of mobile phone should be minimized so that farmers may able to have mobile phone and capable to use it for accessing current information about agribusiness. Illiteracy is another major factor in farmers that pertains the usage of agricultural technology. Therefore, it is necessary that the message/ technology interface should be in local language [38].

III. DATASET DESCRIPTION

The crop recommendation system has been proposed which is based on following parameters:

- N- Ratio of Nitrogen in the soil
- P- Ratio of Phosphorous content in the soil
- K- Ratio of Potassium content in the soil
- Temperature - The temperature in degrees Celsius.
- Humidity- Relative humidity in percentage
- ph - ph value of the soil
- Railfall- Rainfall in mm

The dataset has been downloaded from https://www.kaggle.com. Short summary of dataset is as follows:

- Relation Name: Crop_Recommendation
- # of Instances: 1697
- # of Attributes: 8
- Class Attribute: Label
- # of Class Labels: 15

Table I. Features Description of dataset

S. No	Features Vector	Uniqueness	Unique Value (%)	Distinctness	Attributes
1	N	9	1	136	Numeric
2	P	2	0	117	Numeric
3	K	0	0	73	Numeric
4	temp.	1491	88	1594	Numeric
5	humidity	1453	86	1575	Numeric
6	ph	1431	84	1564	Numeric
7	rainfall	1485	88	1588	Numeric
8	label	0	0	15	Nominal

Table II. Statistical Values of dataset

S.No.	Feature Vector	Min	Max	Mean	S D
1	N	0	140	52.647	38.537
2	P	5	145	58.126	34.164
3	K	5	205	52.031	57.069
4	temp.	8.826	41.949	24.712	4.91
5	humidity	14.258	94.964	65.849	24.433
6	ph	3.505	9.935	6.439	0.87
7	rainfall	5.315	298.56	99.365	50.436

IV. COMPARISON OF CLASSIFICATION MODEL

Statistical analysis and pre-processing of dataset have been done. These processes help analyst to know deep information related to dataset. Boundary data points, outlier's data points, missing data values etc. can be easily identified using these methods. By using specific methods (if necessary) of data pre-processing; data can be refined. This refined dataset can further be used for data analysis. In this research, we have proposed to use five DM classification algorithms which are NB, SVM (SMO), NN (IBk), Decision Tree (PART), and Random Forest. Based on these classification algorithms five classification models has been develop. The comparison of different classification models has been done which are presented graphs.

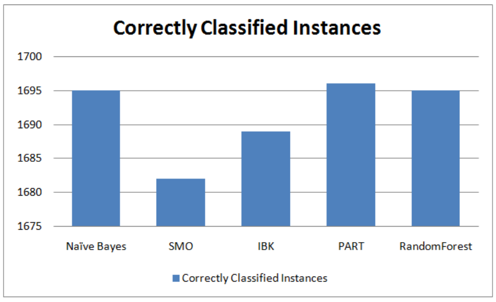

Figure 1: Comparison of Correctly Classified Instances

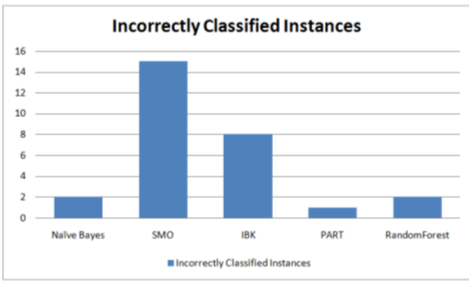

Figure 2: Comparison of incorrectly Classified Instances

© 2025-2027, IJARCS All Rights Reserved

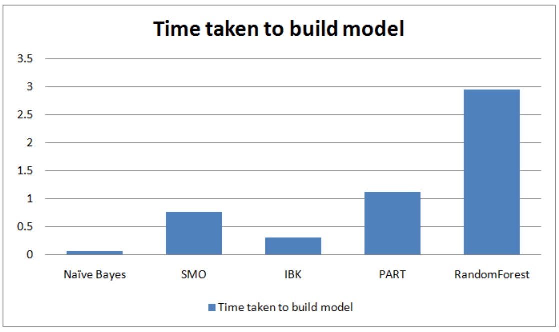

Figure 3: Comparison of Time taken to build model

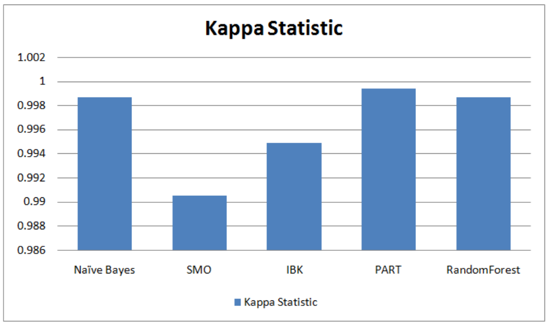

Figure 4: Comparison of Kappa Statistic

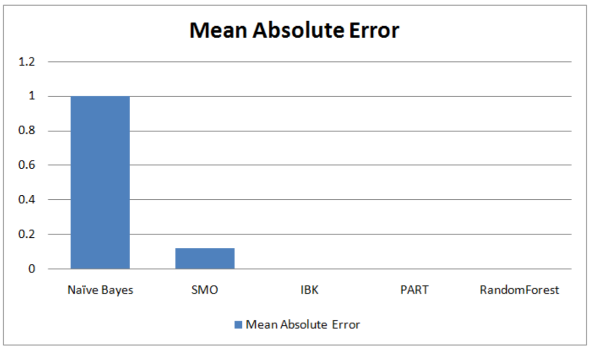

Figure 5: Comparison of MAE

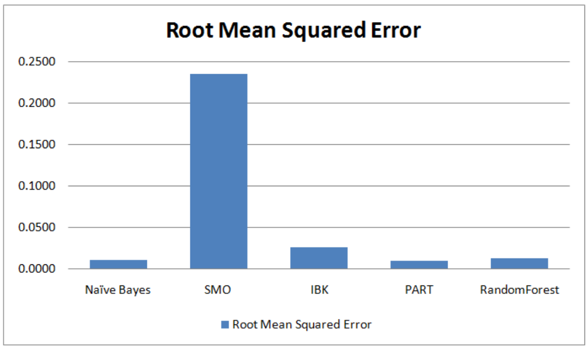

Figure 6: Comparison of RMSE

Table III. Comparison of Classification Models

Classification Algo.	Correctly Classified Instances		Incorrectly Classified Instances	
	Num	in %	Num	in %
NB	1695	99.8821	2	0.1179
SMO	1682	99.1161	15	0.8839
IBK	1689	99.5286	8	0.4714
PART	1696	99.9411	1	0.0589
RF	1695	99.8821	2	0.1179

V. Result and Discussion

Five classification algorithms have been used to analyze dataset. These five DM classification algorithms are NB, SVM (SMO), NN(IBK), Decision Tree (PART), and Random Forest. Precision of developed classification model was NB (99.8821%), SMO (99.1161%), IBK (99.5286%), PART (99.9411%) and Random Forest (99.8821%). Recall of classification model was NB (0.1179%), SMO (0.8839%), IBK (0.4714%), PART (0.0589%) and Random Forest (0.1179%). PART classifier has given the highest accuracy as compare to other classifier also it has minimum number of incorrectly classified instances. The PART classifier has also minimum value for MAE and RMSE which makes it more accurate. On other hands SMO classifier is not suitable for this selected dataset because it has maximum number of incorrectly classified instances with maximum error rates. In this research it has been found that the accuracy of used classification algorithms is more than 99%. But on the basis of comparison of all parameters; we able to conclude that PART algorithm is the best classifier to classify selected dataset.

VI. Acknowledgment

I would like thanks to data owner to provide data as open data for research purpose.

VII. Conclusion

E-agriculture is a new era of agriculture sector. It is just adoption of ICT tools and techniques in agriculture practices. Use of web services, IoT, software, mobile applications etc. are part of this. Processes involved in e-agriculture directly or indirectly generate various databases. These dataset can be utilized by data scientist / analyst for generating effective/ efficient knowledge patterns that may be utilized by decision makers. This study presents a review of literature done by different authors. Experiments related to classification techniques have also been done. Contents which have been discussed has included the application of DM, ML, AI in different sub fields of agriculture like disease detection, yield prediction, crop quality and so on. Classification algorithms have been used to classify different types of crops. For this task five classification algorithms NB, SVM, NN, DT and RF have been selected. Comparison of different parameters such as Accuracy, Kappa Statistic, RMSE, etc. of developed classification model has been presented. Accuracy of classification model was NB (99.8821%), SMO (99.1161%), IBK (99.5286%), PART (99.9411%) and Random Forest (99.8821%). It has been found that decision tree (PART) algorithm is more suitable than other classifier on selected dataset. Integration of new parameters with dataset may generate new results which will be done in future.

VIII. References

[1] Thayer AW, Vargas A, Castellanos AA, Lafon CW, McCarl BA, Roelke DL, Winemiller KO, Lacher TE. Integrating agriculture and ecosystems to find suitable adaptations to climate change. Climate. 2020 Jan 9;8(1):10.

[2] Nassani AA, Awan U, Zaman K, Hyder S, Aldakhil AM, Abro MM. Management of natural resources and material pricing: Global evidence. Resources Policy. 2019 Dec 1;64:101500.

[3] Conrad Z, Niles MT, Neher DA, Roy ED, Tichenor NE, Jahns L. Relationship between food waste, diet quality, and environmental sustainability. PloS one. 2018 Apr 18;13(4):e0195405.

© 2025-2027, IJARCS All Rights Reserved

[4] Benos L, Bechar A, Bochtis D. Safety and ergonomics in human-robot interactive agricultural operations. Biosystems Engineering. 2020 Dec 1;200:55-72.

[5] Lampridi MG, Sørensen CG, Bochtis D. Agricultural sustainability: A review of concepts and methods. Sustainability. 2019 Sep 19;11(18):5120.

[6] Zecca F. The Use of Internet of Things for the Sustainability of the Agricultural Sector: The Case of Climate Smart Agriculture. International Journal of Civil Engineering and Technology. 2019 Sep 18;10(3).

[7] Sonka S. Big data: fueling the next evolution of agricultural innovation. Journal of Innovation Management. 2016 May 4;4(1):114-36.

[8] Meng T, Jing X, Yan Z, Pedrycz W. A survey on machine learning for data fusion. Information Fusion. 2020 May 1;57:115-29.

[9] Evstatiev BI, Gabrovska-Evstatieva KG. A review on the methods for big data analysis in agriculture. InIOP conference series: materials science and engineering 2021 (Vol. 1032, No. 1, p. 012053). IOP Publishing.

[10] Helm JM, Swiergosz AM, Haeberle HS, Karnuta JM, Schaffer JL, Krebs VE, Spitzer AI, Ramkumar PN. Machine learning and artificial intelligence: definitions, applications, and future directions. Current reviews in musculoskeletal medicine. 2020 Feb;13:69-76.

[11] Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: A review. Sensors. 2018 Aug 14;18(8):2674.

[12] Onunga, J O, Nambiro A. Cloud Computing and Other ICT advancements Use in Kenya’s Agricultural Sector: Global Journal of Computer Science and Technology. 2022;22(B1); 49-55.

[13] Umar MA, Sani BM, Suleiman U, Tijjani ML. An Overview of Machine and Deep Learning Technologies Application in Agriculture: Opportunities and Challenges in Nigeria. SLU J. Sci. Technol. 2022;4:84-96.

[14] Wu Y, Ma W. Rural Workplace Sustainable Development of Smart Rural Governance Workplace Platform for Efficient Enterprise Performances. Journal of Environmental and Public Health. 2022;2022(1):1588638.

[15] Rakhra M, Sanober S, Quadri NN, Verma N, Ray S, Asenso E. Implementing Machine Learning for Smart Farming to Forecast Farmers’ Interest in Hiring Equipment. Journal of Food Quality. 2022;2022(1):4721547.

[16] Singh DK, Sobti R, Kumar Malik P, Shrestha S, Singh PK, Ghafoor KZ. IoT-Driven Model for Weather and Soil Conditions Based on Precision Irrigation Using Machine Learning. Security and Communication Networks. 2022;2022(1):7283975.

[17] Songol M, Awuor F, Maake B. Adoption of artificial intelligence in agriculture in the developing nations: a review. Journal of Language, Technology & Entrepreneurship in Africa. 2021;12(2):208-29.

[18] Kumar I, Rawat J, Mohd N, Husain S. Opportunities of artificial intelligence and machine learning in the food industry. Journal of Food Quality. 2021;2021(1):4535567.

[19] Ben Ayed R, Hanana M. Artificial intelligence to improve the food and agriculture sector. Journal of Food Quality. 2021;2021(1):5584754.

[20] Singh RK, Berkvens R, Weyn M. AgriFusion: An architecture for IoT and emerging technologies based on a precision agriculture survey. IEEE Access. 2021 Sep 29;9:136253-83.

[21] Cai Y, Guan K, Lobell D, Potgieter AB, Wang S, Peng J, Xu T, Asseng S, Zhang Y, You L, Peng B. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agricultural and forest meteorology. 2019 Aug 15;274:144-59.

[22] Durga NK, Anuradha G. Plant disease identification using SVM and ANN algorithms. International Journal of Recent Technology and Engineering (IJRTE). 2019 Feb;7(5S4).

[23] Hossain E, Hossain MF, Rahaman MA. A color and texture based approach for the detection and classification of plant leaf disease using KNN classifier. In2019 international conference on electrical, computer and communication engineering (ECCE) 2019 Feb 7 (pp. 1-6). IEEE.

[24] Maimaitijiang M, Sagan V, Sidike P, Hartling S, Esposito F, Fritschi FB. Soybean yield prediction from UAV using multimodal data fusion and deep learning. Remote sensing of environment. 2020 Feb 1;237:111599.

[25] Goap A, Sharma D, Shukla AK, Krishna CR. An IoT based smart irrigation management system using Machine learning and open source technologies. Computers and electronics in agriculture. 2018 Dec 1;155:41-9.

[26] Ferentinos KP. Deep learning models for plant disease detection and diagnosis. Computers and electronics in agriculture. 2018 Feb 1;145:311-8.

[27] Ramesh S, Hebbar R, Niveditha M, Pooja R, Shashank N, Vinod PV. Plant disease detection using machine learning. In2018 International conference on design innovations for 3Cs compute communicate control (ICDI3C) 2018 Apr 25 (pp. 41-45). IEEE.

[28] Dhakal A, Shakya S. Image-based plant disease detection with deep learning. International Journal of Computer Trends and Technology. 2018 Jul;61(1):26-9.

[29] Sindhu S, Sindhu D. Information dissemination using computer and communication technologies for improving agriculture productivity. Development. 2017 Nov;6(6).

[30] Kumar S, Mishra S, Khanna P. Precision sugarcane monitoring using SVM classifier. Procedia Computer Science. 2017 Jan 1;122:881-7.

[31] Matei O, Rusu T, Petrovan A, Mihuţ G. A data mining system for real time soil moisture prediction. Procedia Engineering. 2017 Jan 1;181:837-44.

[32] Rajeswari SK, Suthendran K, Rajakumar K. A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics. In2017 international conference on intelligent computing and control (I2C2) 2017 Jun 23 (pp. 1-5). IEEE.

[33] Ramos PJ, Prieto FA, Montoya EC, Oliveros CE. Automatic fruit count on coffee branches using computer vision. Computers and Electronics in Agriculture. 2017 May 1;137:9-22.

[34] Sellam V, Poovammal E. Prediction of crop yield using regression analysis. Indian Journal of Science and Technology. 2016 Oct;9(38):1-5.

[35] Grinblat GL, Uzal LC, Larese MG, Granitto PM. Deep learning for plant identification using vein morphological patterns. Computers and electronics in agriculture. 2016 Sep 1;127:418-24.

[36] González Sánchez A, Frausto Solís J, Ojeda Bustamante W. Predictive ability of machine learning methods for massive crop yield prediction.

[37] Amatya S, Karkee M, Gongal A, Zhang Q, Whiting MD. Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting. Biosystems engineering. 2016 Jun 1;146:3-15.

[38] Thankachan S, Kirubakaran S. A survey conducted on E-Agriculture with Indian Farmers. International Journal of Computer Science and Mobile Computing. 2014 Feb;3(2):8-14.

© 2025-2027, IJARCS All Rights Reserved