How to Cite:
Balasaranya, K., Ezhumalai, P., Lalluprasad, B., Logesh, B. N., & Reddy, L. I. (2022). Crop cultivation and sustainability for agricultural lands based on soil and atmosphere conditions using modified deep valley algorithm. *International Journal of Health Sciences*, *6*(S2), 12516–12524. https://doi.org/10.53730/ijhs.v6nS2.8304

Crop cultivation and sustainability for agricultural lands based on soil and atmosphere conditions using modified deep valley algorithm

Mrs. K. Balasaranya
Assistant Professor, Department of CSE, R.M.D. Engineering College

Dr. P. Ezhumalai
Professor and Head, Department of CSE, R.M.D Engineering College

Mr. B. Lalluprasad
Project Engineer, Wipro Limited

Mr. B. N. Logesh
Program Analyst Trainee, Cognizant Technology Solutions India Private Limited
Mr. L. Indhiresh Reddy
Student, Department of CSE, R.M.D. Engineering College.

***Abstract*---**Agriculture is the major aspect of human being worldwide. The production of the raw materials for both food and industries may not be sufficient in future given that population growth is rapid. The agriculture techniques was modernised in the past in many ways by shuffling the crop type, introduction of new farming, techniques, fertilizer, pesticides and farming equipments , looking forward to the future ecosystem control technologies can help increase production . It is not only that agriculture raw materials is a need but it also a good source of economy and if it becomes a good stable business it might attract more people into agriculture, increasing the farmer's count which has been falling. The agricultural ecosystem can be sustained by predicting the type of the crop that can be sourced and would give major returns on the yield, this can be done using data science technique and machine will be trained with dataset. This will aid the farmers and vendors for increasing the productivity.

***Keywords*---**crop cultivation, agricultural lands, atmosphere conditions, soil.

Manuscript submitted: 09 March 2022, Manuscript revised: 18 April 2022, Accepted for publication: 27 May 2022

Introduction

Farming is an important part that should be performed in every region in order to make an ecosystem sustainable not only it makes people's life stable but also economical status. Many agricultural products are raw materials for many industries. Apparently agriculture is the very foundation of human life to sustain both in nourishment and modernity. Information is the most valuable thing which helps people to move on further, information related to agriculture can be gathered to observe the pattern of occurrence the process called data mining. This process is used by companies to increase their margin by studying the previous occurrence. Data mining in agriculture includes pre process of data, removing outlying values and detecting mining values and drawing relationship of objects.

Data mining techniques are employed on previous occurrence of weather, temperature, crop cultivation and crop productivity can be increased. Decision support system (DSS) must be implemented to prevent any over heads of decision and support farmers with which crop can be cultivated, DSS is a software system used by analyst to analyse business model and raw dataset to produce useful information and help to predict the feasible option. This system helps farmers by providing suggestive options over regular crop cultivation method which may have repeated crop cultivation.

Supervised technique is the process of training machine with list of inputs and outputs with all the information the machine will be trained to perfection on producing exact prediction when a new input is provided. The machine will be trained with lots of inputs say U and output V and the output V is the derivation of U with relation between different objects being drawn V=f(U), the objective is to map function approximation so that when a new data (U) is given it will be capable of producing an output V. There are two types of machine learning one is classification and the other one is regression which is used in the project, logistic regression, multiclass classification, decision trees and vector machines are few examples of supervised machine learning technique algorithms.

Literature Review

Douglas K et.al., [1] in their research work incorporated data from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS). County level data from the US Department of Agriculture were used. Empirical models anticipated maize and soybean production in the Central region of United States. They also investigated MODIS for its excellence to capture annual variations in yields. In comparison to the commonly used Normalized Difference Vegetation Factor Index, their findings revealed that the two-band Enhanced Vegetation Factor Index (EVI2) from MODIS is a superior predictor of maize yields (NDVI). Addition of agricultural phonology data from MODIS increased model performance. Surprisingly, identifying agricultural regions with intermediate spatial resolution data had little effect on model outcomes when compared with usage of high spatial resolution data.

Department of Agriculture is focussing prediction of crop condition, sustainability and production values at the region and county level, according to Paul C

Doraiswamy et al., [2]. Department of Agriculture's National Agricultural Statistical Service (NASS) of US launches field relevant interviews with selective farm operators. Crop cuttings are predicted to estimate crop yields. National Agricultural Statistics Service (NASS) requires additional geographical data for providing reliable information on crop status and productivity. EPIC (Erosion Productivity Impact Calculator) crop prediction scheme was introduced for regional scale sustainability. The use of remotely sensed data from satellites to measure the amplitude and fluctuation of crop condition metrics in real time is investigated in this study.

This examination was place in the south eastern section of North Dakota, in the semi-arid region. The main goal was to test a strategy for incorporating satellite imagery-derived data into crop productivity. Results are used to predict wheat yields during spring season. Input attributes collected from remote data is used to sense geographic integrity. Real time model calibration simulates attributes throughout the season to guarantee that the predicted and observed circumstances were in agreement. The interface between the satellite data and the crop model was given by SAIL (Scattered by Arbitrary Inclined Leaves), a radioactive transfer. Attributes were simulated on geographic information system based grid structure. It acts as the platform for yield aggregation.

Agriculture is responsible for the majority of the nation's economic contribution around the world. However, due to a lack of ecosystem control methods, the majority of agricultural fields are still underdeveloped. Crop production is not improving as a result it has an impact on the farm economy. As a result, based on plant yield prediction, the development of agricultural production is boosted in the work of D.Sabareeswaran et.al.[3]. Initially, various features such as plant photos, soil parameters, and weather conditions are obtained, and the Firefly (FF) optimization technique for Feature Selection is proposed (FFFS). The most ideal features are then identified using the Modified Fuzzy Cognitive Map (MFCM) algorithm for predicting plant yield growth. The expected outcome is communicated to farmers via smart phones, which aids in identifying plant development and enhancing harvesting. The experimental results suggest that the proposed technique's effectiveness can be compared to that of other prediction techniques.

Proposed System

A machine learning model for Crop yield Prediction is developed, by anticipating results in the form of optimal accuracy by comparing supervised algorithms, to potentially replace easily deployable neural network based classifiers.

Explore data

In exploration of raw data, Load data, verify for cleanliness, and then trim and clean the dataset for analysis is carried out. Meticulously document the processes and provide justifications for cleanliness selections. Variable identification exploration data analysis involves

- Importing essential libraries packages
- Analyzing general attributes
- Finding redundant and missing values
- Counter for unique and common data
- Univariate data analysis includes
- Rename, add, or remove data
- To specify data type

Bivariate and multivariate exploration use

- Pairplot, heatmap, and bar chart plot diagram
- Feature engineering method for detecting outliers
- Partitioning the test and training datasets
- Decision tree, Random forest Algorithm, Logistic regression model metrics comparison is carried out. Using comparison, algorithm to anticipate the outcome the best accuracy is achieved.

Advantages of Proposed System

It excels in removing outliers. Irrelevant variables can be easily found out. Outperforms for the pre processing input which is blend of continuous, discrete data, categorical and sequential. Out-of-bag estimate error prediction is good

Implementation

Machine learning necessitates the collection of large amounts of historical data. Data collection has a significant amount of historical and myths. Raw data cannot be used without pre-processing. It is pre-processed, and it is utilized to choose what kind of algorithm to apply with the model. This model has been trained and tested to perform correctly and forecast with few errors. The accuracy of a tuned model is improved by tuning it from time to time.

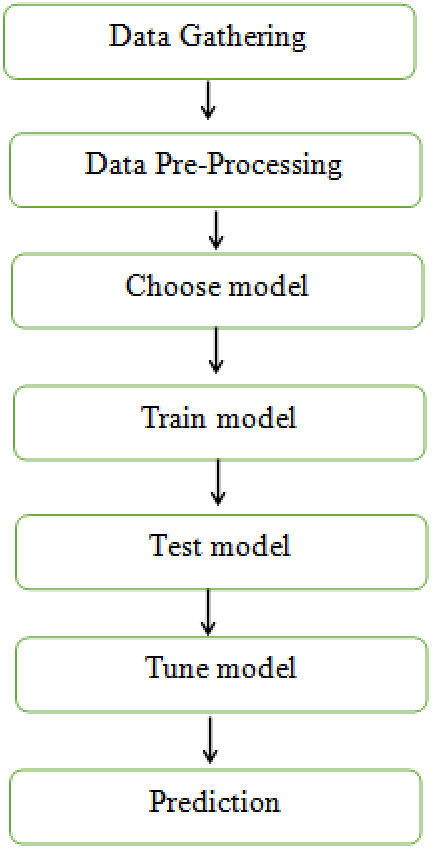

Figure 1: Depiction of Overall Process

Cleaning and Preparation of Data

Importing library packages and loading the specified datasetare carried out. Identifying variables based on shape, type, soil productivity, humidity and evaluating irrelevant values. After training the model that is used to measure model skill when tweaking models. Validation and test datasets are involved while evaluating models. The methods and techniques for cleaning data will differ depending on the dataset.

Predictive Model Construction

Every new incoming information for the model is entered as test data set. Model prediction based on inference concludes after the testing operation on the basis of the training data sets. For crop yield prediction, satellite imagery (Remote Sensing Data) has been frequently employed. Although remote data comprises a large number of energy levels, only a few of them have been used to estimate agricultural yields. However, some people have attempted to generate useful features utilising the bands that are generally overlooked, and they have been successful in enhancing results. Most people rarely look at the high-order moments of the features in this dataset. People have employed techniques such as regression models, Random Forest, and Nearest Neighbor based on these datasets.

Table 1 Dataset Attributes

Attribute	Purpose
Name	Name of the crop
Location	Area of crop cultivation
Region	Region information about cultivation
Cultivation cost	Expense needed for cultivation (per hectare)
Production cost	Expense needed for production (per quintal)
Crop Yield	Total production in terms of Quintal/Hectare
Crop quality	Quality of crop
Soil quality	Quality of soil
Rainfall	Average rainfall recorded in the region
Humidity	Wind and rainfall humidity
Temperature	Mean temperature of the environment
Cost	Production cost of crop yield

Training the Dataset

In the data dataset variable, encapsulate the load data() method. Using the experienced split function, divide the dataset into training and test data. To use any algorithm and train test split class from the sklearn and numpy modules, For example. The feature values are denoted by the X prefix in variable, while the target values are denoted by the y prefix.

- This method randomly divides the dataset in 52:48 ratio (training:testing). Then any algorithm is encapsulated.
- Fit training data into this method in the next line so that the computer may learn from it. The training phase is now complete.

Testing the Dataset

Measurements are available as a new argument in numpy array called 'n' and we want to guess what attributes it is. This is accomplished by utilising the deep valley prediction method, which accepts this array as an input and returns the predicted target benchmark as an output. As a result, the expected target value is 0. Calculation of test score is done as the ratio of correct predictions to overall guesses. It is achieved through scoring technique, which compares the test set's actual values to the projected values.
\

Modified Deep Valley Algorithm

For the reliable crop prediction and its sustainability, a modified deep valley algorithm is implemented. Pseudocode of the algorithm is discussed in the section 5.1.
.

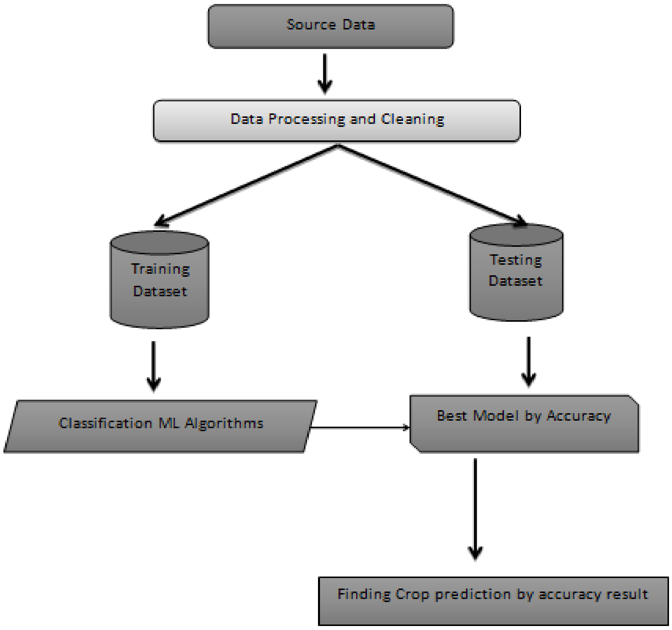

Figure 2: Overall Workflow of the Process

Pseudocode for Modified Deep Valley Algorithm

Random deep valley is formed as clusters and the best cluster and its features are extracted as given below.

```
definition test_attributes:
for each attribute:
find_atmosphere_feature
find_atmosphere_feature(attribute):
for each data:
construct decision tree
find the value valley_peak
calculate the threshold
for each limit:
compare threshold value
form cluster and add in the valley list
for each cluster:
calculate Euclidean distance
find out deep valley peak value
return the best parameters valued cluster
```

Experimental Results

Data visualisations are used to project and demonstrate critical relationships in more visceral and stakeholder-friendly plots and charts than measures of correlation and significance. Data visualisation and exploratory data analysis are fields used to project the output in easier and understandable form.

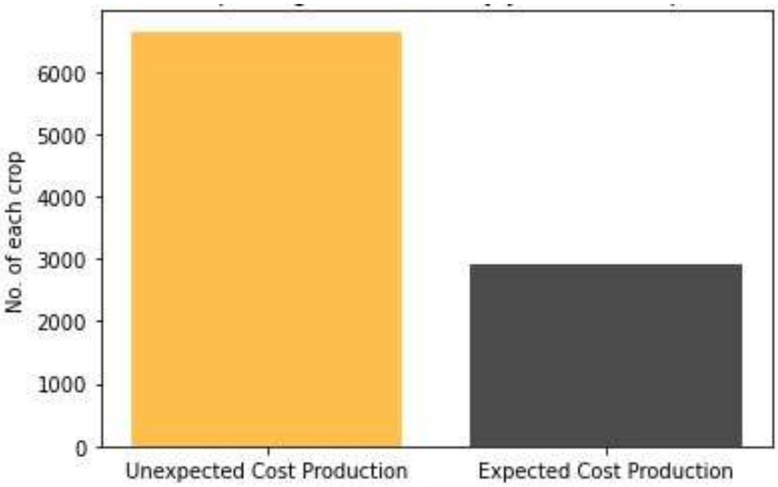

Figure3: Experimental Results for Crop Production Cost

Conclusion

The analytical procedure encompasses data gathering, cleaning, erroneous evaluation, eventual determination, model creation, and evaluation. Deep valley model forecasts the crop fertility using machine learning algorithm with conflicting findings. As a result, the relevant crop assessment insights materialized. Enables farmers to discover about crops that have never been harvested before and see a list of all conceivable crops, which will support them in identifying which crop to cultivate. Furthermore, this methodology leverages formerly collected data, allowing the farmer to obtain awareness about market demand and expenses for chosen crop.

Future Scope

Agricultural department seeks to automate the process of recognising high-yield crops in real time. This method can be automated by displaying the prediction result in a web or by using mobile app or desktop application. Ensemble learning approaches can be utilised to optimise the work in future.

References

[1] Douglas K. Bolton, Mark A. Fried, “Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics”, Agricultural and Forest Meteorology, Volume 173, 15 May 2013, Pages 74-84.

[2] Paul C. Doraiswamy, Sophie Moulin, Paul W. Cook, and Alan Stern, “Crop Yield Assessment from Remote Sensing”, Photogrammetric Engineering and Remote Sensing, DOI:10.14358/PERS.69.6.665, June 2013.
[3] D. Sabareeswaran, R. Gunasundari, “Plant Yield Prediction Model Using Firefly based Feature Selection with Modified Fuzzy Cognitive Maps, International Journal of Pure and Applied Mathematics, Volume 118 No. 5 2018, 339-352.
[4] N. P. Sastra and D. M. Wiharta, “Environmental monitoring as an IoT application in building smart campus of Universitas Udayana”, Proc. Int. Conf. Smart Green Technol. Elect. Inf. Syst. (ICSGTEIS), Oct. 2016, pp. 85–88.
[5] D. Ramesh B. Vardhan "Analysis of crop yield prediction using data mining techniques" International Journal of Research in Engineering and Technology vol. 4 no. 1 pp. 47- 4732015.
[6] O. Elijah, I. Orikumhi, T. A. Rahman, S. A. Babale, and S. I. Orakwue, ―Enabling smart agriculture in Nigeria: Application of IoT and data analytics, Proc. IEEE 3rd Int. Conf. Electro Technol. Nat. Develop. (NIGERCON), Owerri, Nigeria, Nov. 2017, pp. 762–766.
[7] M. S. M. Amin, W. Aimrun, S. M. Eltaib, and C. S. Chan, ―Spatial soil variability mapping using electrical conductivity sensor for precision farming of rice,‖ Int. J. Eng. Technol., vol. 1, no. 1, pp. 47–57, 2004.
[8] E. Husni et al., ―Applied Internet of Things (IoT): Car monitoring system using IBM BlueMix, in Proc. Int. Seminar Intell. Technol. Appl. (ISITIA), Jul. 2016, pp. 417–422.
[9] B. F. Darst, K. C. Malecki, and C. D. Engelman, “Using recursive feature elimination in random forest to account for correlated variables in high dimensional data,” BMC Genet., vol. 19, no. S1, p. 65, Sep. 2018.
[10] J.Y. Hsieh, W. Huang, H.-T. Yang, C.C. Lin, Y.C. Fan, and H. Chen, “Building the rice blast Disease Prediction Model based on Machine Learning and Neural Networks,” Easy Chair World Sci., vol. 1197, pp.1–8, Dec. 2019.
[11] D. H. Zala and M. B. Chaudhri, “Review on use of BAGGING technique in agriculture crop yield prediction,” Int. J. Sci. Res. Develop., vol. 6, no. 8, pp. 675–677, 2018.
[12] F. Shirbani and H. Soltanian Zadeh, “Fast SFFS-based algorithm for feature selection in biomedical datasets,” Amirkabir Int. J. Sci. Res., vol. 45, pp. 43–56, Dec. 2013.
[13] M. Gopal P S and B. R, “Selection of important features for optimizing crop yield prediction,” Int. J. Agricult. Environ. Inf. Syst., vol. 10, no. 3, pp. 54–71, Jul. 2019.