International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

CROP YIELD PREDICTION USING MACHINE LEARNING

THENNARASAN M[1], SURYA L[2], VIGNESH V[3] ,PAVITHRA [4]

[1]*UG Scholar, Department of CSE, Kingston College, Vellore-59*
[2]*UG Scholar, Department of CSE, Kingston College, Vellore-59*
[3]*UG Scholar, Department of CSE, Kingston College, Vellore-59*
[4]Asst.*Professor, Department of CSE, Kingston College, Vellore-59*

Abstract - India is an Agriculture based economy whose the greater part of the GDP comes from cultivating. The inspiration of this undertaking comes from the rising self destruction rates in ranchers which might be because of low collect in crops. Environment and other natural changes have turned into a significant danger in the farming field. AI is a fundamental methodology for accomplishing down to earth and compelling answers for this issue. Foreseeing yield of the harvest from verifiable accessible information like climate, soil, precipitation boundaries and noteworthy harvest yield. We accomplished this utilizing the AI calculation. We did a similar investigation of different AI calculations, i.e., ANN, K Nearest Neighbor, Random Forest, SVM and Linear Regression and picked Random Forest Algorithm which gave a precision of 95%.

In this task a portable application has been created which predicts the harvest yield overall and furthermore for a specific yield. Alongside that, it likewise recommends the client assuming it is the ideal opportunity to utilize the compost or not.

1.INTRODUCTION

IIn India, agribusiness is one of the significant areas as half labor force is engaged with rural exercises. India represents 7.68% of complete worldwide rural result. Gross domestic product of Industry area is $495.62 billion and world position is 12. In Services area, India world position is 11 and GDP is $1185.79 billion. Commitment of Agriculture area in Indian economy is a lot higher than world's normal (6.1%). Yet, Traditional homesteads in India actually have some of least per capita efficiency and rancher incomes[8]. This area likewise require a ton of human endeavors to do different sort of errand like watering crop, developing harvest, spreading pesticides and so on. Soil examination is significant system as it gives supplements present in soil, for example, NPK values and pH esteem. In mechanized soil testing human endeavors will be diminished by observing the nature of soil utilizing soil sensor. Contingent upon the qualities we get from our gadget appropriate rundown of yields is anticipated. Crop expectation is likewise significant boundary to build the yield creation. So based on NPK and pH esteem we foresee the suitable yield alongside the necessary manures so ranchers will defeat the leaving technique drawback.ndia is an Agriculture based economy whose the vast majority of the GDP comes from cultivating. The inspiration of this undertaking comes from the rising self destruction rates in ranchers which might be because of low collect in crops. Environment and other ecological changes have turned into a significant danger in the farming field. AI is a fundamental methodology for accomplishing useful and successful answers for this issue. Anticipating yield of the harvest from authentic accessible information like climate, soil, precipitation boundaries and memorable yield. We accomplished this utilizing the AI calculation. We did a similar investigation of different AI calculations, i.e., ANN, K Nearest Neighbor, Random Forest, SVM and Linear Regression and picked Random Forest Algorithm which gave an exactness of 95%.

In this undertaking a portable application has been created which predicts the harvest yield overall and furthermore for a specific yield. Alongside that, it additionally proposes the client assuming it is the ideal opportunity to utilize the manure or not.

2. Body of Paper

The crop mutation allows the soil to regain the minerals that were used by the crop previously and use the left over minerals for cultivating the new crop. This process will help in maintaining the soil fertility consistently. To know if the soil has reached the point where it is unfit to yield the particular crop, farmer has to experience a loss in yield. One financial year for a farmer is very crucial to accept the loss. The following are the drawbacks of the existing manual System:

- Scope for redundancy
- Time Delay
- Less accuracy
- Needs more human effort
- Requires more laboratory

Table -1: Sample Table format

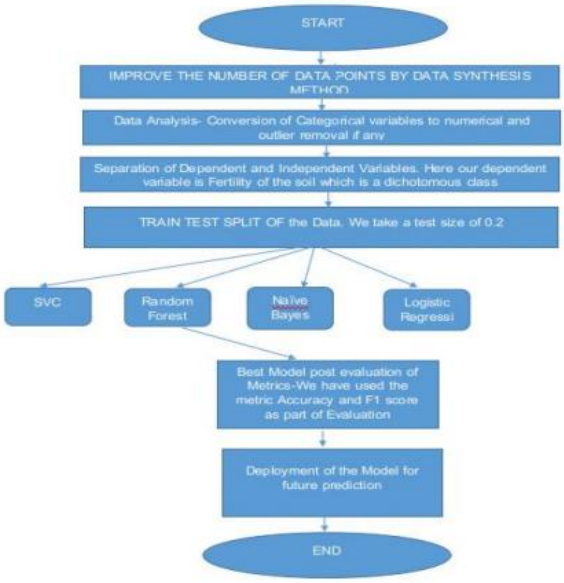


Fig -1: Figure

PROPOSED SYSTEM

Proposed system will determine the fertility of soil by measuring basic parameter effecting soil fertility like PH and electrical conductivity. This will help the farmer in deciding, what crop to bow in the soil, based on the measured nutrient contents present in soil at current state. It will also help in determining what fertilizer to use for the crop. And amount of fertilizer required. Throughout the cultivation process of crop, farmer can test soil fertility n no. of times and take necessary steps to get a good yield. Farmer can maintain the record of his land fertility over the span of time and can access it remotely from anywhere.

Our main objective is the examination, adaptation, and formulation of soil properties and crops growth factors.

The main aim of this project is the examination of macro and micro soil properties such as organic content, essential plant nutrients, that affects the crop yield and find out the rank of a given soil based on the previously graded soil using Supervised Learning and provide the neccessary remedies for the further growth And here we can Able analysis the plant affected disease by using machine learning techniques - random forest

METHODOLOGY:

Basic Process

i. Data Collection**:** Collect the data that the algorithm will learn from.

ii. Data Preparation**:** Format and engineer the data into the optimal format, extracting important features and performing dimensionality reduction.

iii. Training**:** Also known as the fitting stage, this is where the Machine Learning algorithm actually learns by showing it the data that has been collected and prepared.

iv. Evaluation**:** Test the model to see how well it performs.

Tuning**:** Fine tune the model to maximize its performance

EXPLANATION

1. **Data collection**: The dataset used in this project is the data collected from reliable websites and merged to achieve the desired data set. The sources of our datasets are: https://www.kaggle.com/srinivas1/agricuture-crops-production-in-india for crop yielddata. It consists of names of the crops, production, area, average temperature, average rainfall (mm), season, year, name of the states and the districts. Production‘ is the dependent variable or the class variable. There are eight independent variables and 1 dependent variable.
2. **Data Preprocessing**: The purpose of preprocessing is to convert raw data into a form thatfits machine learning. Structured and clean data allows a data scientist to get more precise results from an applied machine learning model. The technique includes data formatting, cleaning, and sampling. Here, data pre-processing focuses on finding the attributes with null values or invalid values and finding the relationships between various attributes as well. Data Pre-processing also helps in finding out the impact of

each parameter on the target parameter. To preprocess our datasets we used EDA methodology. All the invalid and null values were handled by removing that record or giving the default value of that particular attribute based on its importance.

3. **Dataset splitting**: A dataset used for machine learning should be partitioned into two subsets — training and test sets. We split the dataset into two with a split ratio of 80%i.e., in 100 records 80 records were a part of the training set and remaining 20 records

Datasets

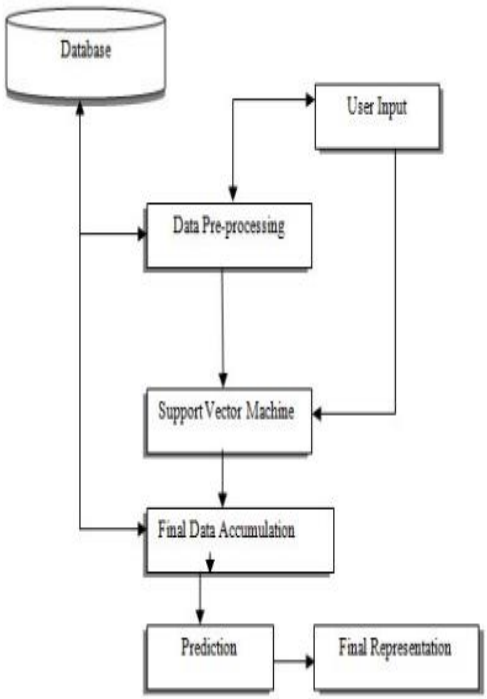
Machine Learning depends heavily on data. It‘s the most crucial aspect that makes algorithm training possible. It uses historical data and information to gain experiences. The better the collection of the dataset, the better will be the accuracy.

The first step is Data Collection. For this project, we require two datasets. One for modelling the yield prediction algorithm and other for predicting weather .i.e. Average Rainfall and Average Temperature. These two parameters are predicted so as to be used as inputs for predicting the crop yield. The sources of our datasets are: _https://en.tutiempo.net/‘ for weather data and _https://www.kaggle.com/srinivas1/agricuture-crops-production-in-india‘ for crop yielddata.

The yield prediction module dataset requires the following columns: State, District, Crop, Season, Average Temperature, Average Rainfall, Soil Type, Area and Production as these are the major factors that crops depend on. _Production‘ is the dependent variable or the class variable. There are eight independent variables and 1 dependent variable. We achieved this by merging the datasets. The datasets were merged taking the location as the common attribute in both. We are considering only two states here, Maharashtra & Karnataka as the suicide rates in farmers in these two States were found to be very high.

SYSTEM ARCHITECTURE DIAGRAM:

System design is the process of the defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements .Systems design could be seen as the application of systems theory to product development. Objectoriented analysis and methods are becoming the most widely used methods for computer systems design. Systems design is therefore the process of defining and developing systems to satisfy specified requirements of the user. The UML has become the standard language in object oriented analysis and design



System architecture is a conceptual model that defines the structure and behaviour of the system. It comprises of the system components and the relationship describing how they work together to implement the overall system.

SYSTEM TEST

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that

program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures : interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

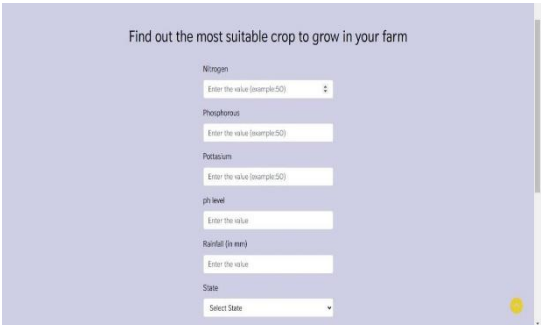


Fig -3: Figure

3. CONCLUSIONS

Agriculture is the backbone of many countries including India. Since integrating the information technology with the agriculture will guide the farmer to improve the productivity. In this proposed work the system described works faster and gives better accuracy in prediction to predict the suitable crops and fertilizers for the field. It includes various parameters of soil to analyse the crop. This prediction makes the farmers to improve the productivity, growth, and quality of the plants

ACKNOWLEDGEMENT

In future this project can be modified by adding few more things like:

- Predicting the pesticide ratio based on the atmospheric and soil parameters for the farm land.

- Database can be created to store the soil details which given by the farmer and also the predicted crop name.

- New CSV file can be created where the new datasets can be added and this can be used for training the model in the future

REFERENCES

1. 1 National Research Council. Alternative Agriculture. National Academy Press. Washington, D.C. 1989.
2. P. Gruhn, F. Goletti, Yudelman M Integrated nutrient management, soil fertility, and sustainable agriculture: current issues and future challenges. International Food Policy Research Institute, Washington, D.C. 2000.
3. A. Mucherino, P. Papajorgji, P.M. Pardalos, A survey of data mining techniques applied to agriculture. Oper. Res. 9 (2009), 121-140.
4. K. Lamorski, Y. Pachepsky, C. Slawin ´ski, R.T. Walczak, Using support vector machines to develop pedotransfer functions for water retention of soils in Poland. Soil Sci. Soc. Amer. J. 72 (2008), 1243-1247.
5. M. Kovacevic, B. Bajat , B. Gajic, Soil type classification and estimation of soil properties using support vector machines, Geoderma. 154 (2009), 340-347.
6. Y. Liu, H. Wang, H. Zhang, and K. Liber, A comprehensive support vector machine-based classification model for soil quality assessment, Soil Tillage Res. 155 (2016), 19-26.
7. Y. Huang, Y. Lan, S.J. Thomson, A. Fang, W.C. Hoffmann, R.E. Lacey, Development of soft computing and applications in agricultural and biological engineering, Comput. Electron. Agric. 71 (2010), 107-127.
8. W. Wu, A.D. Li, X.H. He, et al. A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. Comput. Electron. Agric. 144 (2018), 86-93
9. H. Li, W. Leng, Y. Zhou, F. Chen, Z. Xiu, D. Yang, Evaluation Models for Soil Nutrient Based on Support Vector Machine and Artificial Neural Networks, Sci. World J. 2014 (2014), 478569.
10. M.H. Yang, D.Y. Xu, S.C. Chen, et al. Evaluation of machine learning approaches to predict soil organic matter and pH using Vis-NIR spectra. Sensors. 19(2) (2019), 263.
11. M. Hosseini, S. Rajabi Agereh, Y. Khaledian, et al. Comparison of multiple statistical techniques to predict soil phosphorus. Appl. Soil Ecol. 114 (2017), 123-131.
12. K. Liakos, P. Busato, D. Moshou, S. Pearson, D. Bochtis, Machine Learning in Agriculture: A Review, Sensors. 18 (2018), 2674.
13. J.R. Romero, P.F. Roncallo, P.C. Akkiraju, et al. Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires. Comput. Electron. Agric. 96 (2013), 173-179.
14. M. Sirsat, E. Cernadas, M. Fern ´andez-Delgado, M. Khan, Classification of agricultural soil parameters in India, Comput. Electron. Agric. 135 (2017), 269-279.
15. M.S. Sirsat, E. Cernadas, M. Fern ´andez-Delgado, S. Barro, Automatic prediction of village-wise soil fertility for several nutrients in India using a wide range of regression methods, Comput. Electron. Agric. 154 (2018), 120-133