

# CROPCAST: FORECASTING AGRICULTURE YIELDS

## AUTHORS:

1 Pulimamidi Meghana Reddy,2 Bachireddy Subhashini ,3 Edigi Sai Kiran Goud

B-Tech,4th year students Department of Data Science(CSE)Sphoorthy Engineering College

1meghanareddy22.pmr@gmail.com,2  
3saikirangoudprince@gmail.com

subhashinireddie@gmail.com,

Under the Supervision of Mr Raghavendra (Assistant Professor)

rmichangi@gmail.com

## ABSTRACT:

Agriculture plays a dominant role within the growth of the country's economy. Climate and alternative environmental changes has become a serious threat within the agriculture field. Machine learning (ML) is an important approach for achieving sensible and effective solutions for this drawback. Crop Yield Prediction involves predicting yield of the crop from obtainable historical obtainable information like weather parameter, soil parameter and historic crop yield. This paper target predicting the yield of the crop supported the present information by Random Forest classifier. Real information was used for building the models and also the models were tested with samples. The prediction can help the farmer to predict the yield of the crop before cultivating onto the agriculture fields.

## KEYWORDS:

Machine learning, Crop Yield Prediction, Weather parameters, Soil Parameter, Random Forest Classifier.

## INTRODUCTION:

Agriculture is the backbone of each economy. In an exceedingly country like Bharat, that has ever increasing demand for food thanks to rising population, advances in agriculture sector are needed to satisfy the requirements. From the ancient period, agriculture is taken into account because the main and therefore the foremost culture practiced in Bharat. Ancient folks cultivate the crops in their own land then they

need been accommodated to their wants. Therefore, the natural crops area unit cultivated and is utilized by several creatures like personalities, animals and birds. The greenish merchandise created within the lands that are taken by the creature results in a healthy and welfare life. Since the invention of recent innovative technologies and techniques the agriculture field is slowly degrading.

Due to these, torrential invention folks area unit been targeting cultivating artificial product that's hybrid products wherever there results in associate unhealthy life. Nowadays, fashionable folks don't have awareness concerning the cultivation of the crops in an exceedingly right time and at a right place. Owing to these cultivating techniques the seasonal climatic conditions also are being modified against the elemental assets like soil, water and air that cause insecurity of food. By analyzing of these problems and issues like weather, temperature and a number of other factors, there is no correct answer and technologies to beat things baby-faced by United States of America. Data processing additionally helpful for predicting the crop yield production. Generally, method is that the process of analyzing knowledge from completely different views and summarizing it into helpful data. Data mining software system is associate analytical tool that enables users to research knowledge from many various dimensions or angles, categorize, and summarize the relationships known. Technically, method is

that the process of finding correlations or patterns among dozens of fields in giant relative databases. The patterns, associations, or relationships among all this knowledge will give data. Data will be born-again into data concerning historical patterns and future trends. As an example, outline data concerning crop production will facilitate the farmers determine the crop losses and forestall it in future. Crop yield prediction is a very important agricultural problem. Each and every farmer is usually tries to grasp, what quantity yield can get from his expectation. In the past, yield prediction was calculated by analyzing farmer's previous expertise on a specific crop. The Agricultural yield is primarily depends on atmospheric condition, pests and designing of harvest operation.

#### **LITERATURE SURVEY:**

Crop Yield Prediction Using Machine Learning Algorithm

#### **AUTHORS :**

D.Jayanarayana Reddy Research Scholar Department of Computer Science And Engineering Jawaharlal Nehru Technological University, Dr M. Rudra Kumar Professor and Head, Department of Computer Science and Engineering,

The paper identifies the problems faced by existing methods, such as high cost, low accuracy, and non-linearity, and proposes some objectives for future research, such as selecting optimal features, using deep learning models, and considering fertilizer and irrigation factors. This has a systematic review of the literature on crop yield prediction using machine learning algorithms and also to suggests some directions the suitable crop for the farmers. The paper acknowledges some limitations as it does not provide a clear research question or objective for the systematic.

Crop Cast :

Harvesting the future with interfused machine learning and advanced stacking ensemble for precise crop prediction

#### **AUTHORS:**

Chetan Raju a Department of CSE, JSS Academy of Technical Education , Ashoka D.VDepartment of ISE, JSS Academy of Technical Education , Ajay Prakash B.V c Department of AI and ML, Raja Rajeswari College of Engineering.

The paper proposes a novel method for improving crop prediction accuracy using IMLASE and also this paper claims that this method outperforms several existing techniques(AP-SIM,MSER,DNN,ML,EL) in terms of accuracy, error rate and processing time , and also it proposed that the IML-ASE model consists of three layers : The first layer uses various ensemble techniques as base learners , the second layer acts as meta learner , and the third layer serves as a fine learner for accurate crop prediction . This model leverages information about agricultural , environmental and soil conditions to help farmers make knowledgeable decisions about crop selection. The objective of this paper is to introduce a novel approach for crop prediction using Interfused Machine Learning with Advanced Stacking Ensemble model (IML-ASE) and to develop the IML-ASE model that leverages information about agricultural, environmental, and soil conditions within Agro-Ecological zones to help farmers make informed decisions about crop selection.

#### **EXISTING SYSTEM :**

The agriculture sector needs a huge up gradation in order to survive the changing conditions of Indian economy. Along with the advances in machines and technologies used in farming, useful and accurate information about different matters also plays a significant role in it. This information is being gathered by the use of remote sensors, satellite images, surveys etc. This information along with the knowledge of

subject experts and researchers should be readily available to the farmers in order to exploit its potential worth.

### DISADVANTAGES OF EXISTING SYSTEM :

Less amount of accuracy score

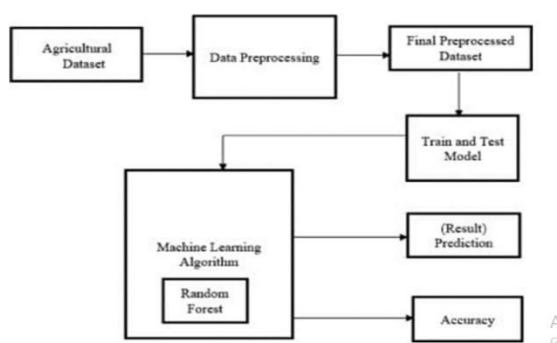
λ Small level data-set

λ Applicable on small level prediction workλ

### PROPOSED SYSTEM :

The agriculture sector needs a huge up gradation in order to survive the changing conditions of Indian economy. Along with the advances in machines and technologies used in farming, useful and accurate information about different matters also plays a significant role in it. This information is being gathered by the use of remote sensors, satellite images, surveys etc. This information along with the knowledge of subject experts and researchers should be readily available to the farmers in order to exploit its potential worth.

### SYSTEM ARCHITECTURE :



### SYSTEM STUDY FEASIBILITY STUDY :

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility

analysis, some understanding of the major requirements for the system is essential.

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

### ECONOMICAL FEASIBILITY:

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

### TECHNICAL FEASIBILITY:

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system. **SOCIAL FEASIBILITY:**

The aspect of study is to check the level of The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it

as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

**SYSTEM DESIGN :**

The Unified Modeling Language (UML) is used to specify, visualize, modify, construct and document the artifacts of an object-oriented software intensive system under development. UML offers a standard way to visualize a system's architectural blueprints, including elements such as: Actors business processesλ (logical) componentsλ activitiesλ

**USE CASE DIAGRAM:**

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

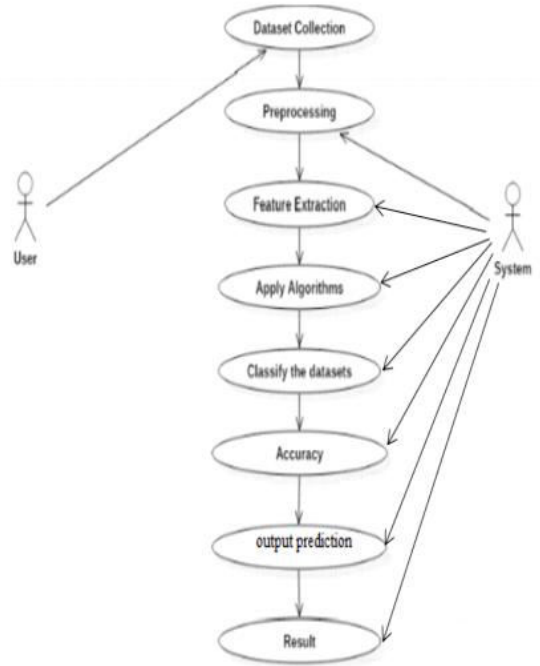
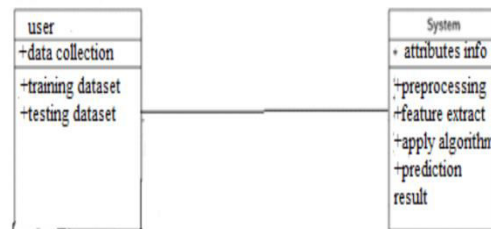


fig1:Usecase diagram

**CLASS DIAGRAM:**

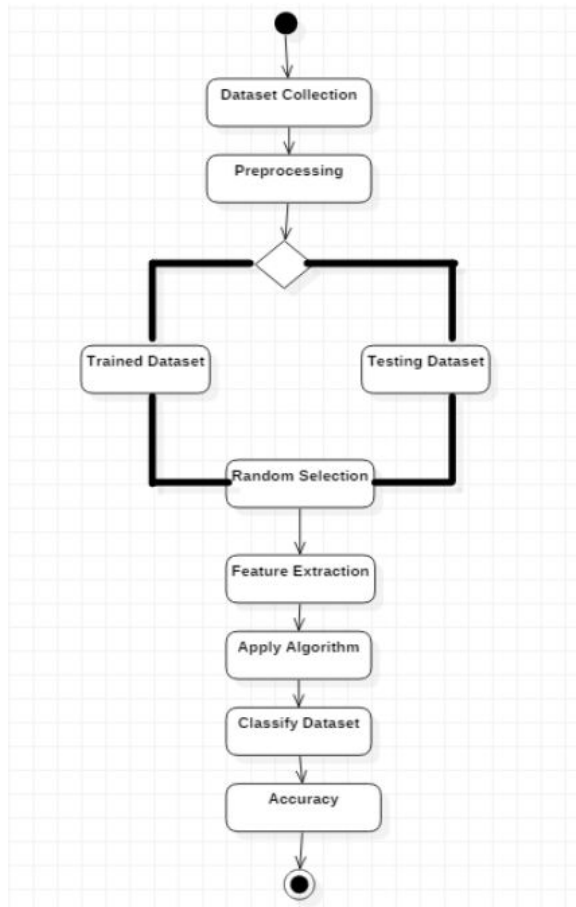
In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



**ACTIVITY DIAGRAM:**

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the

business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.



## IMPLEMENTATION:

1. DATA COLLECTION
2. DATA PRE-PROCESSING
3. FEATURE EXTRATION
4. EVALUATION MODEL

### DATA COLLECTION:

Data collection is a process in which information is gathered from many sources which is later used to develop the machine learning models. The data should be stored in a way that makes sense for problem. In this step the data set is converted into the understandable

format which can be fed into machine learning models. Data used in this paper is a set of cervical cancer data with 15 features . This step is concerned with selecting the subset of all available data that you will be working with. ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer. Data for which you already know the target answer is called labelled data.

### DATA PRE-PROCESSING:

Organize your selected data by formatting, cleaning and sampling from it. Three common data pre-processing steps are: Formatting: The data you have selected may not be in a format that is suitable for you to work with. The data may be in a relational database and you would like it in a flat file, or the data may be in a proprietary file format and you would like it in a relational database or a text file. Cleaning: Cleaning data is the removal or fixing of missing data. There may be data instances that are incomplete and do not carry the data you believe you need to address the problem. These instances may need to be removed. Additionally, there may be sensitive information in some of the attributes and these attributes may need to be anonymized or removed from the data entirely. Sampling: There may be far more selected data available than you need to work with. More data can result in much longer running times for algorithms and larger computational and memory requirements. You can take a smaller representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

### FEATURE EXTRATION :

Next thing is to do Feature extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. The transformed attributes, or features, are linear combinations of the original

attributes. Finally, our models are trained using Classifier algorithm. We use classify module on Natural Language Toolkit library on Python. We use the labelled dataset gathered. The rest of our labelled data will be used to evaluate the models. Some machine learning algorithms were used to classify preprocessed data. The chosen classifiers were Random forest. These algorithms are very popular in text classification tasks.

### **EVALUATION MODEL:**

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can easily generate overoptimistic and over fitted models. There are two methods of evaluating models in data science, Hold-Out and Cross-Validation. To avoid over fitting, both methods use a test set (not seen by the model) to evaluate model performance. Performance of each classification model is estimated base on its averaged. The result will be in the visualized form. Representati

on of classified data in the form of graphs. Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

### **SOFTWARE ENVIRONMENT :**

What is Python : Below are some facts about Python. Python is currently the most widely used multi-purpose, high-level programming language. Python allows programming in Object-Oriented and Procedural paradigms. Python programs generally are smaller than other programming languages like Java. Programmers have to type relatively less and indentation requirement of the language, makes them readable all the time. Python language is being used by almost all tech-giant companies

like – Google, Amazon, Facebook, Instagram, Dropbox, Uber... etc.

### **6.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 .

**CONCLUSION:**

In conclusion, In this project we are performing crops prediction for district level. So main aim is to find the dataset which contains production details of past 10-12 years also details about climatic parameters and soil parameters like rainfall, temperature, moisture, soil contents etc. details. These factors will help in the prediction of the crops by using various classifiers on the given dataset. Thus, various factors actualize both categorical and continuous dependent variables, it is a type of supervised learning algorithm mostly used for classification problems. What this algorithm does is, it splits the population into two or more homogeneous sets based on the most significant attributes making the groups as distinct as possible. The classification algorithm will give us best split on different features for selection of most suitable crop among the population. The feature selection methodology of random forest algorithm made it suitable for prediction of suitable crops.

**REFERENCES:**

[1] J. Wang, H. Yue, and Z. Zhou, "An improved traceability system for food quality assurance and evaluation based on fuzzy classification and neural network," *Food Control*, vol. 79, pp. 363–370, Sep. 2017.

[2] Q. Lin, H. Wang, X. Pei, and J. Wang, "Food safety traceability system based on blockchain and EPCIS," *IEEE Access*, vol. 7, pp. 20698–20707, 2019. [3] G. Alfian, M. Syafrudin, U. Farooq, M. R. Ma'arif, M. A. Syaekhoni, N. L. Fitriyani, J. Lee, and J. Rhee, "Improving efficiency of RFID-based traceability system for perishable food by utilizing IoT sensors and machine learning model," *Food Control*, vol. 110, Apr. 2020, Art. no. 107016.

[4] F. Liu, Y. Wang, Y. Jia, S. Hu, L. Tu, and C. Tang, "The egg traceability system based on the video capture and wireless networking technology," *Int. J. Sensor Netw.*, vol. 17, no. 4, pp. 211–216, Apr. 2015.

[5] X. Xiao, Z. Fu, L. Qi, T. Mira, and X. Zhang, "Development and evaluation of an intelligent traceability system for frozen tilapia fillet processing," *J. Sci. Food Agricult.*, vol. 95, no. 13, pp. 2693–2703, Oct. 2015. [6] M. M. Aung and Y. S. Chang, "Traceability in a food supply chain: Safety and quality perspectives," *Food Control*, vol. 39, pp. 172–184, May 2014.

[7] E. Abad, F. Palacio, M. Nuin, A. G. D. Zárate, A. Juarros, J. M. Gómez, and S. Marco, "RFID smart tag for traceability and cold chain monitoring of foods: demonstration in an intercontinental fresh fish logistic chain," *J. Food eng.*, Vol. 93, no. 4, pp. 394–399, aug. 2009. [8] F. Tian, "A supply chain traceability system for food safety based on HACCP, blockchain & internet of things," in *proc. Int. Conf. Service syst. Service manage.*, Jun. 2017, pp. 1–6.

[9] J. F. Galvez, J. C. Mejuto, and J. Simal-Gandara, "Future challenges on the use of blockchain for food traceability analysis," *TrAC Trends Anal. Chem.*, vol. 107, pp. 222–232, Oct. 2018.

[10] G. P. Danezis, A. S. Tsagkaris, F. Camin, V. Brusica, and C. A. Georgiou, "Food authentication: Techniques, trends & emerging approaches," *TrAC Trends Anal. Chem.*, vol. 85, pp. 123–132, Dec. 2016.

[11] M. Thibaud, H. Chi, W. Zhou, and S. Piramuthu, "Internet of Things (IoT) in high-risk environment, health and safety (EHS) industries: A comprehensive review," *Decis. Support Syst.*, vol. 108, pp. 79–95, Apr. 2018 [12] K. Salah, N. Nizamuddin, R. Jayaraman, and M. Omar, "Blockchain-based soybean traceability in agricultural supply chain," *IEEE Access*, vol. 7, pp. 73295–73305, Aug. 2019.

[13] T. Bosona and G. Gebresenbet, "food traceability as an integral part of logistics management in food and agricultural supply chain," *food control*, vol. 33, no. 1, pp. 32–48, sep. 2013. [14] M. Lv, Y. Liu, J. Geng, X. Kou, Z. Xin, and D. Yang, "Engineering

nanomaterials-based biosensors for food safety detection,” *Biosensors Bioelectron.*, vol. 106, pp. 122–128, May 2018. [15] K. Behnke and M. Janssen, “Boundary conditions for traceability in food supply chains using blockchain technology,” *Int. J. Inf. Manage.*, vol. 52, Jun. 2020, Art. no. 101969 [16] R.-Y. Chen, “An intelligent value stream-based approach to collaboration of food traceability cyber physical system by fog computing,” *Food Control*, vol. 71, pp. 124–136, Jan. 2017.