

Crop Prediction Using Various Feature Selection Techniques and Classifiers Based on Agricultural Environment Characteristics

L. Shalini¹, Buddupu Joseph Srinivas Reddy^{2*}, Sai Vardhan Chakali³, Chandan Kumar Chaubey⁴, David Chandra⁵

¹Professor, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, India

^{2,3,4,5}Student, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, India

Abstract: An important role that agriculture plays in strengthening the US economy is as follows. The advancement of humanity can be attributed to agriculture. India is an agricultural nation in the United States of America, with a fruit-based economy. The foundation of everything in the United States is agriculture. In agricultural businesses, crop selection is vital. The yield's commitment might be entirely dependent on a number of variables, such as production stages, market costs, and national legislation. Numerous agricultural changes are necessary to enhance the financial system in India. Agricultural productivity may be increased by using technology to learn techniques that are effectively used in the field. Precious and accurate data on a wide range of issues play a major role in agriculture, in addition to all the things within the sphere of machinery and inventions utilized in it. The purpose of the gadget is to implement a yield dedication strategy, a random woodland, and a regressor tree judgment in order to enable this selection to address several challenges related to agriculture and farmers. This is strengthening the Indian economy and producing more plants.

Keywords: crop yield prediction, lasso, kernel ridge, enet, stacked regression.

1. Introduction

A sharing of success stories in the field of generation enhancement will help builders understand and develop their capabilities. The reality that valuable and strategic data is shared either openly or privately among agronomists is the foundation of the statistics trade. An open mindset among agronomists is linked to their desire to percentage facts. The extent and degree of record changes are determined by this open connection. We leverage internet technologies like HTML and CSS to construct online utilities. We generate a dataset by gathering data from several sources, integrating it into a website that forecasted harvest prices, and analysing the results using a non-linear analysis. They were eventually found on the 0 crop and priority list. Post data in our application and share this information with agronomists, whose data is gathered We generate a dataset by gathering data from several sources, integrating it into a website that forecasted harvest prices, and analysing the results using a non-linear analysis. They were

eventually found on the 0 crop and priority list. Post data in our application and share this information with agronomists, whose data is gathered and kept on a MySQL server. We send current information via the program.

2. Literature Survey

[1] Rice Yield Prediction Using Support Vector Machines, 2019. Support vector machines (SVMs) are now widely used in the world of computer software. Because they can reproduce, they are often employed inside the provision. The development of SVM-primarily based category fashions for Indian rice yield prediction is the focus of this work. The kernel polynomial function of SVM education, okay-cross validation, and the multiple classification technique have all been used in experiments. The Department of Economics and Statistics of the Indian Department of Agriculture provided data on rice output in India for this study. The four-triple cross-validation method produced a great predicted accuracy of 75.06% for the four-year relative average growth. MATLAB software is utilized in this work's experiments.

[2] Method for selecting crops to maximise yield using machine learning techniques, 2015. The protection of American agriculture's food supply and economic growth are significantly aided by agricultural coverage. One of the main challenges in agricultural planning is selecting the crop or crops to plant. It is dependent on a number of factors, such as government coverage, market pricing, and production pace. Numerous scholars have examined crop forecasting, climate forecasting, soil classification, and crop class in order to inform agricultural planning through the application of statistical techniques or system learning approaches. Crop selection becomes a conundrum when there are multiple options for when to plant and use available land. In order to solve the crop selection problem, optimize crop internet at a certain time in the yield duration, and ultimately obtain the greatest financial increase of the location, this study employs a technique known as the Crop Selection Method (CSM). The suggested method has the potential to boost internet yields.

*Corresponding author: josephreddy6601@gmail.com

[3] Machine Learning Crop Data System, 2020. As far as we know, India is the second most famous country in the United States. The majority of people in India and around the world work in agriculture. Farmers repeatedly plant the same plants without experimenting with other crop varieties, and they apply fertilizers in unknown amounts without knowing enough about the type and quantity of the fertilizer. As a result, it simultaneously affects yield, causes soil acidity, and damages the top layer. As a result, we developed a system that uses device learning algorithms to raise farmers' productivity levels. Based solely on soil composition and meteorological factors, our method will suggest the greatest crop that is appropriate for a certain soil.

The gadget also provides information on the type and quantity of fertilizers needed for seed cultivation. Therefore, farmers may cultivate novel plant varieties, increase yields, and prevent soil pollution by using our equipment.

[4] Give an overview of the usage of machine gaining knowledge of strategies, 2019. In India, one of the most important yet lowest paid jobs are agriculture. By cultivating the most desirable crops, machine learning could change the profitability situation and spur an increase in agriculture. The goal of this paper is to provide a yield prediction through the application of several system study techniques. The mean absolute errors are used to compare the overall performance of those approaches. Farmers will be able to choose which crop to grow in order to maximize yield with the help of machine learning algorithms that consider variables such as area, temperature, rainfall, and more.

[5] A Machine Learning Approach to Predicting Yields and Success Rates, 2019. About 23% of India's GDP comes from agriculture, which also employs 59% of the country's workforce. It is the second-biggest wild crop in India. Technology has the potential to help farmers produce more. Offering a variety of plants might assist a farmer in choosing which crop to grow. The primary focus of the research is to use regression neural community modelling to forecast the yields of different plants. Records of the harvest cycle are kept for the entire year, including the summer, kharif, rabbi, and autumn. The Government of India website is where the dataset was obtained.

Crop region, fruit, nation, region, season, 12 months, and production or yield were the experimental characteristics taken into consideration for the study, which ran from 1998 to 2014. Two lakh forty thousand statistics are included in the statistics set. Pandas and Pandas profiling tools in Python are used to filter the dataset and extract data from the state of Maharashtra. The version that used a multilayer perceptron neural network was evolved. It was first obtained with a 45% accuracy inside the RMS optimizer. Later, by employing more layers, modifying the weight, biasing, and converting the Adam optimizer, the accuracy was increased to 90%. This study details the creation of a different yield prediction version utilizing an ANN and a three-manner neural network. In order to build up a model for return prediction, an ANN model creates components to validate the relationship between the use of many input and output samples. Activation characteristic: The

activation process makes use of a linear rectifier (Rely) block. Techniques for propagation and backpropagation are employed.

[6] Application of machine gaining knowledge of methods in agricultural crop production, 2016. In an attempt to reduce the amount of research on the applicability of device studying methods in the field of agricultural production, this article was prepared. Methods/Statistical Analysis: This technique is a novel approach to product manufacturing. Precise and timely crop production forecasts are crucial for critical decisions made by the Department of Economics and Statistics, such as import-export, pricing, advertising distribution, and so forth. It should be noted, nonetheless, that these first assessments are not objective opinions since they need a number of descriptive evaluations, some of which may be based on a wide range of qualitatively different elements. As a result, an objective statistically valid crop output prediction has to be expanded. There is a vast quantity of information available because to the advancements in computer and statistics storage. Conclusions: The endeavour of extracting information from this complex raw data has resulted in the development of state-of-the-art methodologies and procedures as well as system knowledge that can be applied to combine yield estimate with information technology. By using these novel methodologies to many database variables, this study aims to evaluate them and enable the discovery of significant correlations. Use or enhancement: There are several methods that use artificial neural networks, such as Bayesian community theory, informative networks, decision trees, and regression assessment. Agriculture is the topic of time series analysis, Markov chain model, okay-way clustering, and okay closest and neighbour vector machines.

[7] Machine Learning: Applications in Indian Agriculture, 2016. The foundation of a developing financial system, such as India's, is agriculture. The majority of people rely on agriculture as their source of income. In India, agriculture has become a risky profession due to resource depletion, limited land areas, rising labour costs and production costs, and unpredictability surrounding weather patterns and other variables. Modern advancements ought to be applied to all fields and have already produced notable gains in several areas. However, agriculture stopped making significant advancements.

Agrarian work is necessary for the Indian economy. A relatively recent area of computer technology that can be effectively used to agriculture is machine learning. This can simplify the process of modernizing the industry's traditional training techniques while keeping costs low. This newsletter's goal is to broaden farmers' horizons by listing and assessing the numerous gadgets learning packages available in Indian agriculture and helping them take on more work at different levels.

A. Objective

- Compiling statistics from several sources.
- Methods for cleaning and evaluating data are employed to transform raw statistics into technique records.
- A crop value update machine that operates quickly is created by evaluating the collected data using a system

learning device.

- By using an ensemble of classifiers, the model gains efficiency and robustness.
- A mission's set of tactics aids in our ability to make wise decisions.
- Development of an online application for user registration and data collection.

B. Existing System

Using common sense, Chawla, I. et al. (2019, August) anticipated how time collecting styles will affect record usage. Precipitation and temperature were two of the forecasting parameters they examined. They state "excellent yield" in their forecast type with "excellent yield." In order to improve the prediction's efficiency, Chaudhary, A. N. et al. (2018, August) used three algorithms: the clustered means, the previous set of rules, and the Bayesian algorithm. They then hybridized the algorithm, considering factors like location, rainfall, soil type, and the device's ability to determine which crop is appropriate to grow based on the indicated characteristics.

Disadvantages of Existing System:

- There is poor performance. He paused longer.
- The state-of-the-art method considers just positive plants—all other vegetation and characteristics are excluded.
- Quite a slow construction.
- Interpreting is challenging. A priceless computation.
- System research may result in redundancy, which is a drawback.

3. Proposed System

The suggested induction device would advance agricultural technology in the United States of America and help farmers select crops that will provide larger yields. It may also be used to increase yields to increase agricultural capital and decrease farmer losses. Therefore, the suggested machine will help to lessen the issues that farmers face and stop them from attempting suicide. It will also serve as a resource to give farmers the useful records they need to obtain excessive yields, maximizing yields, which in turn reduces farmers' mortality. Cost and lessen challenges. Monitoring the productivity of different types of vegetation may help maximize production. It also helps choose the right crop for selected fields and seasons, which alleviates problems for farmers in the agricultural region. As a result, the suggested gadget offers a method for forecasting agricultural output. Before growing the sphere to generate a better yield, a farmer would verify the crop output in step with acre.

Advantages of Proposed System:

- Beneficial for people who live far from cities.
- Enhancing time utilization performance. The drop happens gradually.
- A strong and safe apparatus.
- In the sector of agriculture, this machine is helpful since it allows us to choose suitable plants, from which the circumstances and selection determine the price,

among many other benefits.

- Because machine learning allows us to generate predictions only based on data, it eliminates the uncertainty and headaches associated with using large and intricate trouble spaces.

Proposed Algorithm:

- Random Forest Algorithm
- Decision Tree Regression Algorithm

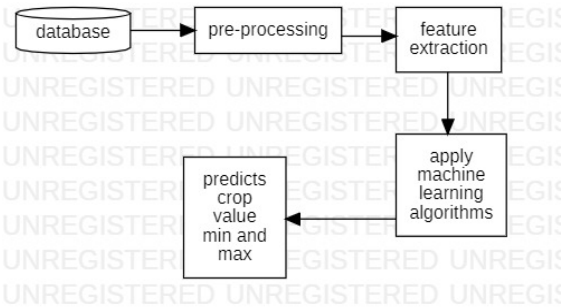


Fig. 1. System architecture

System Requirements:

Hardware Requirements:

- System: Pentium-IV
- Speed: 2.4GHZ
- Hard disk: 40GB
- Monitor: 15VGA color
- RAM: 512MB

Software Requirements:

- Operating System: Windows XP
- Coding language: Python
- IDE: Flask web app

Software Environment:

Python:

- Python is an interpretive, interactive, literal item-oriented language with a high degree of abstraction. Python's architecture makes it simple to investigate. It has fewer syntactic structures than other languages and typically uses English key phrases whereas other languages utilize punctuation.
- Python is interpreted; during runtime, Python is handled by an interpreter. It is not necessary to complete this system before starting to execute. It is similar to PHP and PERL.
- Python is interactive; you may use the interpreter to write packages right away while sitting at a command line.
- Python is object-oriented: Python supports an item-oriented programming paradigm in which objects contain code.
- Python is a beginner's language. For inexperienced programmers, Python is an amazing language that can be used to enhance a broad range of applications, including video games, web browsers, and word processors.

Modules:

- Data Collection
- Calculate yield of production
- Predict crop value
- Accuracy on test set

Data Collection:

- This is the first serious step in really developing a computer that learns about models and collects data. This is a crucial stage that establishes the version's level of precision: our model performs better the more and better information we have.
- There are many other ways to obtain information, such as manual intervention, net feed, and many more.
- This Indian yield angle dataset is taken from another source.

Calculate yield of production:

- In this challenge, the harvest's price is determined by the harvest's quality, which is determined by order. It is recommended to use the minimum and maximum yield using this process.

Predict crop value:

- Crop values are forecasted in this module by applying device learning algorithms to the collected and trained data. such that, depending on entry, we can get the lowest and highest value for each given neighbourhood.

Accuracy on test set:

- He got 90.7% accuracy at the test set.

Data Flow Diagram:

- A bubble chart is another name for a DFD. It is a simple graphical formalism that may be used to represent a system in terms of the data that goes into the device, the various analyses that are done on that data, and the outputs that are produced by it.
- One of the essential modelling tools is data flow diagrams (DFDs). It is employed to simulate system components. These elements include the machine ways, the data that the technique uses, the external device that matches the device, and the data flows inside the system.
- The DFD illustrates the flow of information through the system and how modifications are made to it over time. The information may be moved from input to output using this graphical method that shows the records' flow and the alterations that can be made.
- Another name for a DFD is a bubble chart. At every level of abstraction, a machine may be represented using a DFD. Layers inside a DFD comprise incremental data glide and man-or-woman actions.

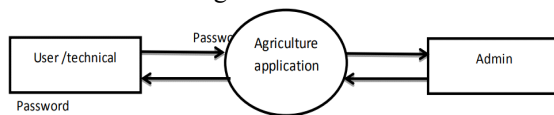


Fig. 2. LEVEL-0

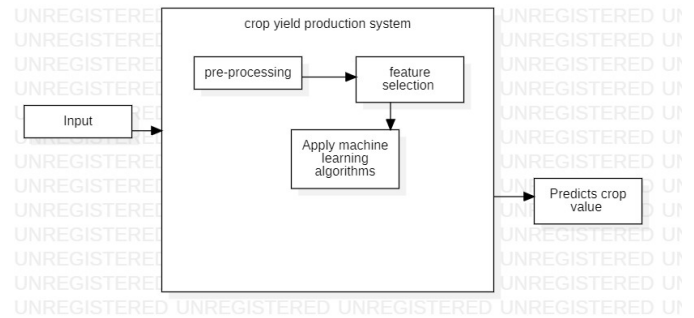


Fig. 3. UML diagram

Any alternative visual representation of the General Semantic Model is called UML. A comprehensive syntax for managing an object's whole development life is provided by UML.

Advantages:

- Show how to employ item-oriented concepts to represent entire structures, not just segments of programs.
- Clearly define the link between executable code and standards.
- Take into consideration the scale components included in intricate and important buildings.
- To develop a language format that is helpful to both computers and humans.
- UML specifies a variety of styles for system representation.
- The class model assumes a robust form.
- The dynamic behaviour of devices is expressed in a kingdom version.

A "person use case" is a type of use case. Business version scenarios and message flows are included.

- An example denotes the application of work units
- The deployment model provides information for distribution plans.

Use Case Diagram:

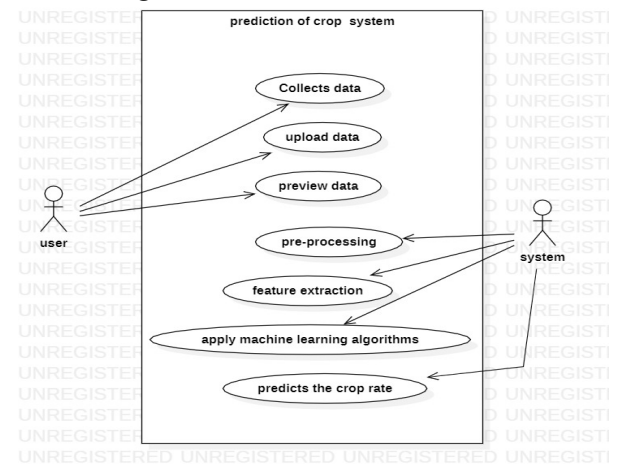


Fig. 4.

The specifications needed to operate the device are displayed in the use case diagrams. Although they are helpful for showing management and/or current buyers, use cases provide a far

higher premium for true improvement since they capture the "sense" of genuine needs. A use case is shown as a horizontal ellipse and specifies a set of sports that are valuable in some way to the interest.

Activity Diagram:

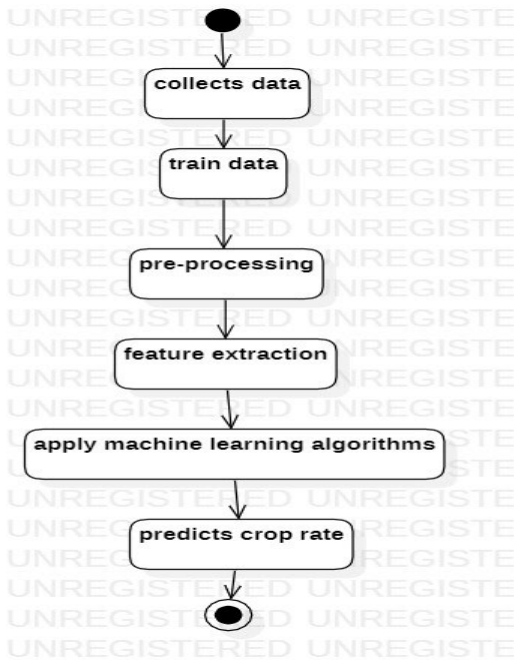


Fig. 5.

A activity diagram is a visual representation of a methodical, interest-based process that aids in decision-making, release management, and concurrency. An activity diagram can be used to explain how a device's components operate and gradually wear out. A begin node, an interest quit node, and intermediate activities make up an activity diagram.

Sequence Diagram:

Commonly used for both assessment and layout, visually models the glide of good judgment on your device so you can verify both your documentation and your good judgment. The most often used dynamic UML modelling technique for specifying device behaviour is a collection diagram.

Input Design and Output Design:

Input Design:

The conduit between the user and the facts machine is the enter method. In order to transform transactional facts into a form that can be used in a process, it entails improving specifications and methods for information preparation. These can be done by computer programs that analyse data from printed or written scripts, or they can be done manually. The humans will help with the introduction of the keys. promptly categorized into flaws. Controlling the quantity of input needed, preventing errors, averting delays, preventing additional procedures, and maintaining system simplicity are the specializations of input planning.

The login process is intended to be secure and easy while protecting user privacy. The committee's suggestions became these:

- Which documents must be provided for input?

- How are the documents encoded or arranged?
- A different box to aid staff in entering data.
- Techniques for preparing entry validation and responding to errors.

Objectives:

- The process of transforming an input description into a laptop computer is known as input design. This process is essential for preventing errors in the information input process and for determining the best path for the management to take in order to obtain the appropriate data from the automated device.
- To handle vast volumes of information, this is accomplished by building the proper record entry shelves. The enter technique was created to eliminate mistakes and streamline data access. The purpose of this records access screen is to enable the completion of all fact operations. It also offers a way to see the information.

Information is far verified for correctness as it is input. Monitors may be used to enter data. The user is provided with necessary instructions as needed, so they are not in an immediate kingdom. Therefore, the enter layout's purpose is to design an enter layout that is simple to understand.

Output Design:

- It's a good product that provides the information honestly and satisfies the needs of the giver. Any device's outputs are used to communicate the procedure's results to users and other systems. The output strategy outlines the movement of records for both printed output and quick request. For the user, it is the most reliable and up-to-date source of statistics. The connection device's reasonable and effective output design gets better, assisting the user in making judgments.
- The development of computer goods has to be well-planned and structured; the best outputs need to be created with every output detail ready so that users may operate the device effectively and without trouble. Selecting the exact output that satisfies the requirements requires careful analysis of the computer's output.
- Decide how to deliver the information.
- Construct a document, report, or other layout with the data that the device produced.

One or more of the following functions should be supported by the information device's output arrangement.

- Disseminate information on topics other than sports, current celebrity, or predictions
- The outcome; critical events, opportunities, queries, or prompts.
- Get the motion going.
- Verify the activity.

System Study:

Feasibility Study:

This includes an analysis of the mission's viability and the provision of an opinion along with a job outline and cost

estimates. It is crucial to conduct a feasibility study of the suggested device as part of the system evaluation. This is to make sure that the employer isn't always burdened by the suggested equipment. Some knowledge of the essential machine requirements is needed for feasibility studies.

System Testing:

The goal of testing out is to identify mistakes. The process of testing involves identifying any potential flaw or vulnerability in the paintings that are currently being completed. This offers a means of determining the functioning of individual parts, assemblies, subassemblies, and/or completed work. There is a method for testing software.

The software system no longer fails in an inappropriate way and satisfies the needs and expectations of the users. Certain types of tests exist. Every check type is in line with a certain examination need.

4. Conclusion

The extent and quality of the alternating information are determined by this open dating. an extensive assessment of technologies to accurately increase agricultural yields via innovation. This project suggests a novel, clever device for

forecasting product expenses. Using the fleet to forecast is the main concept. When several classifiers are used, the prediction is higher than when only a few classifiers are used. Furthermore, a grading system was used to determine the layout of the categorization outcomes. This tool is intended to project the crop's future charge price.

References

- [1] Manpreet Kaur, Heena Gulati, Harish Kundra, "Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications", *International Journal of Computer Applications*, vol. 99, no. 12, August 2014.
- [2] J. Meng, "Research on the cost of agricultural products circulation and its control under the new normal economic development," *Commercial Times*, no. 23, pp. 145-147, 2016.
- [3] A. Kaloxylos et al., "Farm management systems and the future Internet era," *Comput. Electron. Agricult.*, vol. 89, pp. 130-144, Nov. 2012.
- [4] L. Nannan, L. Tongsheng, Y. Zhengsong, R. Yang, M. Yuanyuan, and L. Yongsheng., "Factors influencing farmers' adoption of new technology based on Logistic-ISM model-a case study of potato planting technology in Dingxi City, Gansu Province," *Progress in Geography*, vol. 33, no. 4, pp. 542-551, 2014.
- [5] Y. Wang, "A neural network adaptive control based on rapid learning method and its application," *Advances in Modeling and Analysis*, Vol. 46(3), pp. 27-34, 1994.