



CROP YIELD PREDICTION USING RANDOM FOREST ALGORITHM

Shanmuga Priya M¹, Lin Eby Chandra M², Kumaran M³

Student, Department of CSE, Jaya Engineering College, Chennai, India¹

Assistant Professor, Department of CSE, Jaya Engineering College, Chennai, India²

Professor, Department of CSE, Jaya Engineering College, Chennai, Country India³

Abstract: Most agricultural crops have been badly affected by the effect of global climate change in India. In terms of their output over the past 20 years. It will allow policy makers and farmers to take effective marketing and storage steps to predict crop yields earlier in their harvest. This project will allow farmers to capture the yield of their crops before cultivation in the field of agriculture and thus help them make the necessary decisions. Implementation of such a method with a web-based graphic software that is simple to use and the machine learning algorithm can then be distributed. The results obtained are granted access to the farmer. And yet there are various methods or protocols for such very data analytics in crop yield prediction, and we are able to predict agricultural productivity with guidance of all those algorithms. It utilizes a Random Forest Algorithm. By researching such problems and issues such as weather, temperature, humidity, rainfall, humidity, there are no adequate solutions and inventions to resolve the situation we face. In countries like India, even in the agricultural sector, as there are many types of increasing economic growth. In addition, the processing is useful for forecasting the production of crop yields.

Keywords: predictive modeling, feature selection, data preprocessing, regression analysis, agricultural forecasting, remote sensing data, weather data, soil data, precision agriculture, crop management, decision support system, machine learning models, data – driven farming, yield optimization, agriculture technology, big data in agriculture.

I. INTRODUCTION

India is the land of agriculture and it is the major source of economy. 70% of Indian population directly relies on agriculture. The common problem existing among the young Indian farmers is to choose the right crop based on the location, humidity, temperature, rainfall. Due to this, they face a serious setback in productivity. Our work proposes to help farmers determine the predict crop type and price by doing analysis on its various parameters and to suggest crops based on the results obtained. The system uses the Classification algorithm of Random Forest to improve the efficiency of Crop Recommendation System. The system maps the location, temperature, humidity, pH value and rainfall to predict the list of suitable crops for the soil and it also provides cost of the crop. Hence it leaves upon the user to decide on the crop to be sown. Thus, the system helps to provide knowledge to the dilettante farmers. Agriculture plays a pivotal role in ensuring food security and sustainable development worldwide. The ever-growing global population necessitates increased agricultural productivity and efficient resource management. One of the key challenges in modern agriculture is to accurately predict crop yields, as this knowledge enables farmers and stakeholders to make informed decisions regarding planting, harvesting, and resource allocation.

To address this challenge, the project "Crop Yield Prediction Using Random Forest Algorithm" leverages the power of machine learning and data science. The focus of this project is to develop a robust and accurate predictive model that harnesses the capabilities of the Random Forest algorithm to forecast crop yields. This project marries data-driven technology with agricultural practices, offering a novel approach to enhance farming efficiency, optimize resource utilization, and ultimately contribute to global food security

II. PROPOSED SYSTEM

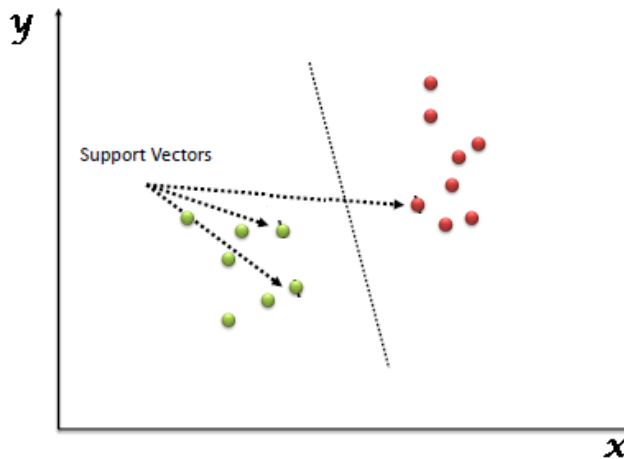
The system prepared predict major crops yield in a particular district in Tamil Nadu. The client on their first login has to register themselves on the application on android phone. Once the user logs into the system he gets all the access for predicting crop yield and using the input such as location, temperature, pH value, rainfall and humidity depends on their forming land environment. After submitting the inputs, it's redirect into Firebase. The firebase is an intermediate between user input and trained data set. The input goes to the trained data, where it processes random forest algorithm to predict crop and price. After the prediction, the predicted value passes to the fire base. That firebase gives the predict value to the user on android application. The proposed system for "Crop Yield Prediction Using Random Forest Algorithm" is designed to provide a comprehensive solution that leverages data analytics and machine learning to accurately predict



crop yields. This system aims to assist farmers and stakeholders in making informed decisions to optimize agricultural productivity, resource allocation, and ultimately contribute to global food security. Here's an overview of the components and functionalities of the proposed system.

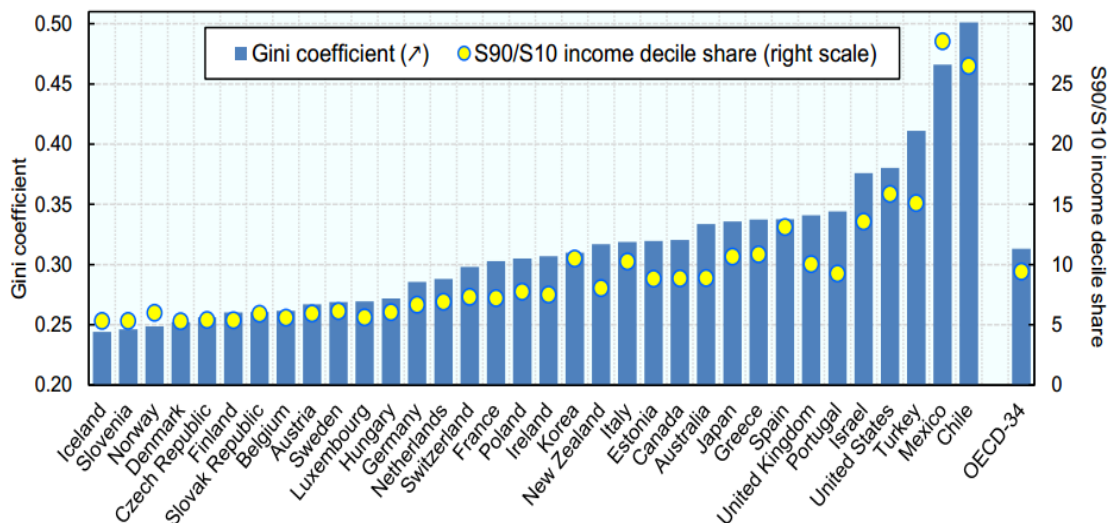
III. METHODOLOGY

Mastering machine learning algorithms isn't a myth at all. Most of the beginners start by learning regression. It is simple to learn and use, but does that solve our purpose? Of course not! Because, you can do so much more than just Regression! Think of machine learning algorithms as an armory packed with axes, sword, blades, bow, dagger etc. You have various tools, but you ought to learn to use them at the right time. As an analogy, think of 'Regression' as a sword capable of slicing and dicing data efficiently, but incapable of dealing with highly complex data. On the contrary, 'Support Vector Machines' is like a sharp knife – it works on smaller datasets, but on them, it can be much stronger and powerful in building models.



Random forest algorithm:

With increase in computational power, we can now choose algorithms which perform very intensive calculations. One such algorithm is "Random Forest", which we will discuss in this article. While the algorithm is very popular in various competitions.



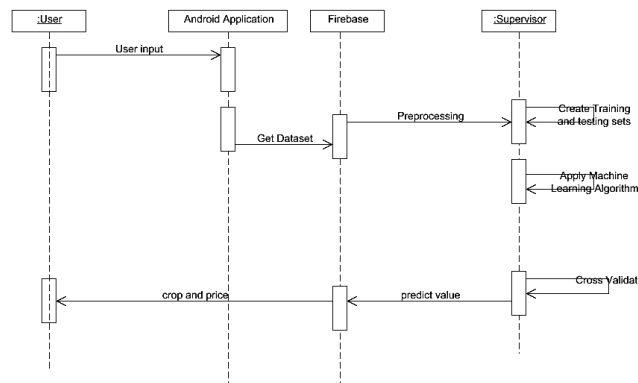
The algorithm of Random Forest

Random forest is like bootstrapping algorithm with Decision tree (CART) model. Say, we have 1000 observation in the complete population with 10 variables. Random forest tries to build multiple CART model with different sample and different initial variables. For instance, it will take a random sample of 100 observations and 5 randomly chosen initial variables to build a CART model. It will repeat the process (say) 10 times and then make a final prediction on each observation. Final prediction is a function of each prediction. This final prediction can simply be the mean of each prediction.



Sequence Diagram:

A Sequence diagram is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of Message Sequence diagrams are sometimes called event diagrams, event sceneries and timing diagram.



Use Case Diagram:

Unified Modeling Language (UML) is a standardized general-purpose modeling language in the field of software engineering. The standard is managed and was created by the Object Management Group. UML includes a set of graphic notation techniques to create visual models of software intensive systems. This language is used to specify, visualize, modify, construct and document the artifacts of an object oriented software intensive system under development.

Dataset Collection:

Our crop prediction project dataset are collected from kaggle.com. Data is pre-processed after collection of various records. The dataset contains a more number of records, where some records are with some missing values. Those missing records have been removed from the dataset and filled the data using python package techniques (Pandas, NumPy).

Implementation:

The classification Algorithms always produces the best results. We are using Random Forest Algorithm to predict the crop and price using ML. On an analysis conducted within various algorithms, the Random Forest was found to provide highest efficiency and precision compared to Decision tree. Because RF contain number of decision tree algorithm, that take the average to improve the predictive accuracy of dataset. Hence the RF algorithm is used in the proposed system to find the suitable crop and crop price.

CODING

Once the design aspect of the system is finalizing the system enters into the coding and testing phase. The coding phase brings the actual system into action by converting the design of the system into the code in a given programming language. Therefore, a good coding style has to be taken whenever changes are required it easily screwed into the system.

INTEGRATION TESTING

Integration testing is a systematic technique for construction the program structure while at the same time conducting tests to uncover errors associated with interfacing. i.e., integration testing is the complete testing of the set of modules which makes up the product. The objective is to take untested modules and build a program structure tester should identify critical modules. Critical modules should be tested as early as possible. One approach is to wait until all the units have passed testing, and then combine them and then tested. This approach is evolved from unstructured testing of small programs. Another strategy is to construct the product in increments of tested units. A small set of modules are integrated together and tested, to which another module is added and tested in combination. And so on. The advantages of this approach are that, interface dispenses can be easily found and corrected.

The major error that was faced during the project is linking error. When all the modules are combined the link is not set properly with all support files. Then we checked out for interconnection and the links. Errors are localized to the new module and its intercommunications. The product development can be staged, and modules integrated in as they complete unit testing. Testing is completed when the last module is integrated and tested.

TESTING

Testing is a process of executing a program with the intent of finding an error. A good test case is one that has a high



probability of finding an as-yet –undiscovered error. A successful test is one that uncovers an as-yet- undiscovered error. System testing is the stage of implementation, which is aimed at ensuring that the system works accurately and efficiently as expected before live operation commences. It verifies that the whole set of programs hang together. System testing requires a test consists of several key activities and steps for run program, string, system and is important in adopting a successful new system. This is the last chance to detect and correct errors before the system is installed for user acceptance testing.

The software testing process commences once the program is created and the documentation and related data structures are designed. Software testing is essential for correcting errors. Otherwise the program or the project is not said to be complete. Software testing is the critical element of software quality assurance and represents the ultimate the review of specification design and coding. Testing is the process of executing the program with the intent of finding the error. A good test case design is one that as a probability of finding an yet undiscovered error. A successful test is one that uncovers a yet undiscovered error.

At the culmination of integration testing, software is completely assembled as a package. Interfacing errors have been uncovered and corrected and a final series of software test-validation testing begins. Validation testing can be defined in many ways, but a simple definition is that validation succeeds when the software functions in manner that is reasonably expected by the customer. Software validation is achieved through a series of black box tests that demonstrate conformity with requirement. After validation test has been conducted, one of two conditions exists.

* The function or performance characteristics confirm to specifications and are accepted.

* A validation from specification is uncovered and a deficiency created.

Deviation or errors discovered at this step in this project is corrected prior to completion of the project with the help of the user by negotiating to establish a method for resolving deficiencies. Thus the proposed system under consideration has been tested by using validation testing and found to be working satisfactorily. Though there were deficiencies in the system they were not catastrophic

IV. CONCLUSION

In conclusion, the project "Crop Yield Prediction Using Random Forest Algorithm" holds immense promise in revolutionizing modern agriculture. By amalgamating cutting-edge data analytics and the power of the Random Forest algorithm, this system emerges as a pivotal tool in the pursuit of enhanced crop yield prediction and sustainable farming practices. The significance of precise crop yield forecasts cannot be overstated, as they not only facilitate resource allocation but also provide farmers with the foresight necessary for effective risk mitigation. With the ability to incorporate diverse data sources, including historical yield records, climate data, and soil quality, the system offers a comprehensive solution for the multifaceted challenges faced by today's agricultural sector. By marrying technology and agriculture, this project underlines the potential of data-driven farming to optimize crop production, minimize environmental impact, and ensure consistent food supplies in the face of growing global demand. The user-friendly interface, integrated recommendations, and data visualization tools make this system accessible and practical for farmers and stakeholders. In this era of agricultural transformation, "Crop Yield Prediction Using Random Forest Algorithm" not only supports the quest for food security but also fosters sustainability, thereby heralding a new frontier in farming practices.

REFERENCES

- [1] . S.Veenadhari, Dr Bharat Misra, Dr CD Singh.2019."Machine learning approach for forecasting crop yield based on climatic parameters.". International Conference on Computer Communication and Informatics (ICCCI), 2014.
- [2] . P.Priya, U.Muthaiah M.Balamurugan.Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology(IJESRT),2018.
- [3] . D Ramesh ,B Vishnu Vardhan. Analysis Of Crop Yield Prediction Using Data Mining Techniques. IJRET: International Journal of Research in Engineering and Technology.(IJERT),2015.
- [4] . F. Baheri, F. Davardoost and V. Ahmadzadeh, "Data mining with learning decision tree and Bayesian network for data replication in Data Grid", Proceedings of the Third International Conference on Contemporary Issues in Computer and Information Sciences, pp. 51- 51, 2012.
- [5] . Abdullah, A., Brobst, S, Pervaiz.I., Umer M.,and A.Nisar.2004. Learning dynamics of pesticide abuse through data mining. Proceedings of Australian Workshop on Data mining and Web Intelligence, New Zealand, January,2004.
- [6] . C.T. Dhanya and D. Nagesh Kumar, "Data mining for evolution of association rules for droughts and floods in India using climate inputs", J. of Geo. Phy. Res.114: 1-15, 2009.
- [7] . VS Boddu, BNK Reddy, MK Kumar, "Low-power and area efficient N-bit parallel processors on a chip," 2016 IEEE annual India conference (INDICON), pp. 1-4, 2016.