Performance Evaluation of Machine Learning Algorithms for Crop Yield Prediction

Veena K[1], Shankar N B[2], Anand Reddy G M[3], Deepika M[4]

Assistant Professor, Department of Computer Science & Engineering[1,4]
Associate Professor, Department of Computer Science & Engineering[2,3]
R.L. Jalappa Institute of Technology, Doddaballapur, Karnataka, India

Abstract: *Agriculture is the backbone of India and also plays an important role in Indian economy by providing a certain percentage of domestic product to ensure the food security. For most developing countries, agriculture is the primary source of revenue. Modern agriculture is a constantly growing approach for agricultural advances and farming techniques. But now-a-days, food production and prediction is getting depleted due to unnatural climatic changes, which will adversely affect the economy of farmers by getting a poor yield and also help the farmers to remain less familiar in forecasting the future crops. This research work helps the beginner farmer in such a way to guide them for sowing the reasonable crops by deploying machine learning, one of the advanced technologies in crop prediction. The modern technologies can change the situation of farmers and decisions making in agricultural field in a better way. Python is used as a front end for analyzing the agricultural data set. Jupyter Notebook is the data mining tool used to predict the crop production. The parameter includes in the dataset are soil nutrient values like Potassium(K), Nitrogen(N), Phosphorous(P) and Temperature, Rainfall, Humidity*

Keywords: Crop Prediction, Food production, Machine Learning Algorithms, Random Forest Algorithm, SVM, Decision Trees.

I. INTRODUCTION

India is ranked 2nd worldwide in farm output. The crop yield of plants relies on different factors like on climatic, geographical, organic, political and financial elements. For farmers, it is difficult when there is more than one crop to grow especially when the market prices are unknown to them. Citing the Wikipedia statistics, the farmer suicide rate in India has ranged between 1.4 and 1.8 per 100000 total population, over a 10-year period through 2005. While 2014 saw 5650 farmer suicides. In recent times, it has become inevitable to use technology to create awareness about cultivation. The seasonal climatic conditions are also being changed against the fundamental assets like soil, water and air which lead to insecurity of food. In a scenario, crop yield rate is falling short of meeting the demand consistently and there is a need for a smart system which can solve the problem of decreasing crop yield. Therefore, to eliminate this problem, we propose a system which will provide crop prediction based on economic and environmental factors to reap the maximum yield out of it for the farmers which will sequentially help meet the elevating demands for the food supplies in the country. The proposed system uses machine learning to make the predictions. The system will provide crop yield and crop selection based on weather attributes suitable for the crop to get the maximum yield out of it for the farmers. The system makes predictions of the productions of crops by studying the factors such as Nitrogen (N), Potassium (K), Phosphorous (P), rainfall, temperature, pH values. The system also helps in suggesting whether a particular time is the right one to use fertilizers.

Crop yield prediction is an important agricultural problem. Every farmer always tries to know how much yield will be produced and whether it meets their expectations. In the past, yield prediction was calculated by analysing a farmer's previous experience on a particular crop. The Agricultural yield is primarily dependent on weather conditions pests and planning of harvest operation. Accurate information about the history of crop yield is an important thing for making decisions related to agricultural risk management.

The system takes the required input from the farmers or sensors such as N, P, K, Temperature, Humidity and pH values. This all inputs data applies to machine learning predictive algorithms like K-Nearest Neighbor (KNN), Random Forest,

DOI: 10.48175/IJARSCT-12758

ISSN
2581-9429
IJARSCT

Decision tree and XG Booster to identify the pattern among data and then process it as per input conditions. The system recommends the crop for the farmers.

II. PROBLEM STATEMENT

Due to the critical challenges in predicting the crops manually, farmers feel more forced to adopt to intensive farming practices in order to increase quality and quantity of the crops".

III. RELATED WORKS

As we know the fact that, India is the second largest population country in the world and majority of people in India have agriculture as their occupation. [1] Farmers are growing same crops repeatedly without trying new variety of crops and they are applying fertilizers in random quantity without knowing the deficient content and quantity. So, this is directly affecting on crop yield and also causes the soil acidification and damages the top layer. So, the authors designed the system using machine learning algorithms for betterment of farmers. In this system they suggest the best suitable crop for particular land based on content and weather parameters and also the designed system provides information about the required content and quantity of fertilizers, required seeds for cultivation. Hence by utilizing this system farmers can cultivate a new variety of crop which may increase in profit margin and can avoid soil pollution.

In general, agriculture is the backbone of India and also plays an important role in Indian economy by providing a certain percentage of domestic product to ensure the food security. [2] But now-a-days, food production and prediction is getting depleted due to unnatural climatic changes, which will adversely affect the economy of farmers by getting a poor yield and also help the farmers to remain less familiar in forecasting the future crops. This research work helps the beginner farmer in such a way to guide them for sowing the reasonable crops by deploying machine learning, one of the advanced technologies in crop prediction. Naive Bayes, a supervised learning algorithm puts forth in the way to achieve it. The seed data of the crops are collected here, with the appropriate parameters like temperature, humidity and moisture content, which helps the crops to achieve a successful growth. In addition as the software, a mobile application for Android is being developed. The users are encouraged to enter parameters like temperature and their location will be taken automatically in this application in order to start the prediction process. In further an android application also developed for the farmers for their convenience where, the farmer are provided with options just to enter the temperature, humidity, soil moisture and pH. Using machine learning the consolidation of data will be getting automatically and the identified crop will be predicted, while the machine learning based crop prediction system has predicted the crop successfully for the given input in reference with the preceding seed data.

[3] Agriculture is one of the fundamental occupations for majority of the countries in the world. Especially, in developing nations like India, the country is primarily driven by agriculture sector, where agriculture and its associated businesses are the backbone of the Economy making it the integral revenue generator. With technological advancements in the recent years, crop yield prediction has gained wide importance, and has shown to have significant impact on the revenue generated from agriculture in every season. Multiple factors influence crop yield prediction, which in turn makes it a non-trivial and challenging task. Despite many proposed works in the area, crop yield prediction lacks a unified solution. This paper brings out the need for a unified framework through a comparative study of standard algorithms and attributes. In this paper considered algorithms are Linear Regression, Random Forest, K Nearest Neighbors (KNN) and Stochastic Gradient Descent (SGD). Our results show that Random Forest outperforms the other standard algorithms by showing 91.62% accuracy in crop yield prediction. Further evaluation is done where the attributes that affect the crop yield most are ranked according to their impact based on Mean Absolute Error (MAE). With this, the authors make a case for the need for a unified approach for crop yield prediction.

Today Agriculture Sector is a major contributor to Indian Economy. [4] In a country like India, which has ever increasing demand of food due to rising population, advances in agriculture sector are required to meet the needs. Therefore Crop Yield Prediction remains a challenging task in this domain. There are various parameters that affect the yield of crop like rainfall, temperature, fertilizers, pesticides, ph level, and other atmospheric conditions and parameters. Accurate yield prediction is required to be done after understanding the functional relationship between yield and these parameters. For this many researchers have applied machine learning algorithms on comprehensive datasets for predicting crop yield.

This paper discusses various machine learning approaches towards crop yield prediction in India. Further in this work, Machine learning approaches have been executed on the agricultural data to evaluate the best performing technique.

[5] Agriculture is one of the important sources of survival and one of the most important factors in the economic growth of the country. Researchers in the field of the agriculture have studied and implemented different well-organized mechanisms which would predict and increase the crop yield sand make agriculture highly profitable. In this paper we have tried to predict the crop yield, suggest the best crop thereby improve the quality and profitability of the agriculture sector by processing huge volume of data often called as Big Data using Hadoop platform. The paper also focuses on the soil type and helps to find which particular crop would be suitable for a particular type of soil. In case of crop yield soil plays an important role and soil information is predicted by considering the weather details of the previous years. Hence the study will predict the suitability of a crop for a particular climatic condition and the possibilities of improving the crops quality by using weather and disease related data sets. Predicting the quality of crops is quite complex and need to take care of multiple approaches. This research paper help to identify mechanisms to get good quality and improved crop yields using a new algorithm named as "Agro algorithm" implemented in Hadoop platform and uses Hadoop framework to handle large amount of data sets.

[6] Agriculture is the backbone of the country. This paper predicts the yield of almost all kinds of crops that are planted in India. This script makes novel by the usage of simple parameters like State, district, season, area and the user can predict the yield of the crop in which year he or she wants to. The paper uses advanced regression techniques like Kernel Ridge, Lasso and ENet algorithms to predict the yield and uses the concept of Stacking Regression for enhancing the algorithms to give a better prediction. In our research, which we found in the previous research papers is that everyone uses climatic factors like rainfall, sunlight and agricultural factors like soil type, nutrients possessed by the soil (Nitrogen, Potassium, etc.) but the problem is we need to gather the data and then a third party does this prediction and then it is explained to the farmer and this takes a lot of effort for the farmer and he doesn't understand the science behind these factors. To make it simple and which can be directly used by the farmer this paper uses simple factors like which state and district is the farmer from, which crop and in what season (as in Kharif, Rabi, etc.). [7] Agriculture and its allied sectors are undoubtedly the largest providers of livelihoods in rural India. The agriculture sector is also a significant contributor factor to the country's Gross Domestic Product (GDP). Blessing to the country is the overwhelming size of the agricultural sector. However, regrettable is the yield per hectare of crops in comparison to international standards. This is one of the possible causes for a higher suicide rate among marginal farmers in India. This paper proposes a viable and user-friendly yield prediction system for the farmers. The proposed system provides connectivity to farmers via a mobile application. GPS helps to identify the user location. The user provides the area & soil type as input. Machine learning algorithms allow choosing the most profitable crop list or predicting the crop yield for a user-selected crop. To predict the crop yield, selected Machine Learning algorithms such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), Multivariate Linear Regression (MLR), and K-Nearest Neighbor (KNN) are used. Among them, the Random Forest showed the best results with 95% accuracy. Additionally, the system also suggests the best time to use the fertilizers to boost up the yield.

[8] Agriculture is a primary supply of Income. It is vital for the survival of the environment. Practically every part of life, people rely on a vast range of agricultural products. Farmers must adapt to climate change while also meeting increased demand for more food of higher quality. To boost agricultural output and growth, farmers must be aware of the weather conditions, which will help them choose the best crop to cultivate under those situations. IoT-based smart farming improves the overall agriculture system by monitoring the field in real time. This system keeps track of a multitude of characteristics, including humidity, temperature, and soil, and displays them in crystal clear real-time. In the agricultural sector, machine learning is utilized to increase crop output and quality. The application of appropriate algorithms to sensed data can aid in crop recommendation. [9] An important issue for the purposes of agricultural planning is a reliable yield estimate for the many crops involved in the planning. Machine learning is an approach to provide practical and efficient solutions to this problem. Many comparisons of ML methods for yield prediction have been made for the most accurate technique. Generally, the number of evaluated crops and techniques is too low and does not provide proper information for agricultural planning purposes. This paper compares the predictive accuracy of ML algorithm for crop yield prediction. People of India are practicing agriculture for years but the results are never satisfying due to various factors that affect the crop yield. To fulfil the needs of around 1.2 billion people, it is very important to have a good yield

of crops. Due to factors like soil type, precipitation, region, seed quality, season, lack of technical facilities etc. The crop yield is directly influenced. Hence, new technologies are necessary for satisfying the growing need and farmers must work smartly by opting new technologies rather than going for trivial methods. In this paper, an Association Rule Mining technique integrating features of the Eclat algorithm and Genetic algorithm into the proposed method. The idea is to use the eclat technique of association rule mining to create rules and to use genetic algorithms to further refine those rules. [10] Agriculture is the largest contributor of the GDP in our country. But still the farmers do not get the exact worth price of the crops. It is mostly due to improper irrigation or imprecise crops selection, sometimes the crop yield is less than expected. In order to add efficiency to the whole agriculture process a "Predicting Agricultural produce using Machine Learning Techniques" is proposed in the following project. This system aims at predicting the agricultural produce by obtaining area on Google Maps API and further calculate area under cultivation.

IV. METHODOLOGY

Machine Learning is undeniably one of the most influential and powerful technologies in today's world. Machine learning is a tool for turning information into knowledge. In the past 50 years, there has been an explosion of data [10]. This mass of data is useless; we analyse it and find the patterns hidden within. Machine learning techniques are used to automatically find the valuable underlying patterns within complex data that we would otherwise struggle to discover. The hidden patterns and knowledge about a problem can be used to predict future events and perform all kinds of complex decision making. To learn the rules governing a phenomenon, machines have to go through a learning process, trying different rules and learning from how well they perform. Hence, why it's known as Machine Learning.

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.

v. Tuning: Fine tune the model to maximize its performance.

V. ALGORITHMS USED

Machine Learning offers a wide range of algorithms to choose from. These are usually divided into classification, regression, clustering and association. Classification and regression algorithms come under supervised learning while clustering and association comes under unsupervised learning.

- Clustering: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behaviour.
- Association: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.
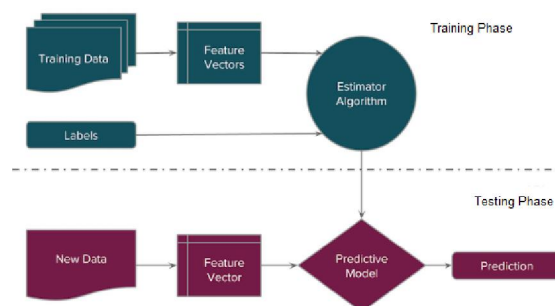
VI. SYSTEM ARCHITECTURE

Figure 4.1: System Architecture

Architecture diagrams can help system designers and developers visualize the high-level, overall structure of their system or application for the purpose of ensuring the system meets their users' needs. They can also be used to describe patterns that are used throughout the design. It's somewhat like a blueprint that can be used as a guide for the convenience of discussing, improving, and following among a team.

VII. Implementation

- Take soil type and area as inputs
- These values are given as input to the random forest implementation in the backend and the corresponding predictions are returned.
- The algorithm returns a particular predicted crop.

VIII. RESULTS

The table 8.1 summarizes the various algorithms that were explored in this work and found that XG Booster provides 99.45 percentage of accuracy.

Table 4.1: Summary of the Approaches

Serial No	Algorithm	Accuracy
1	K Nearest Neighbour	97.89
2	Decision Tree	98.72
3	Random Forest	97.00
4	XG Booster	99.45

KNN algorithm
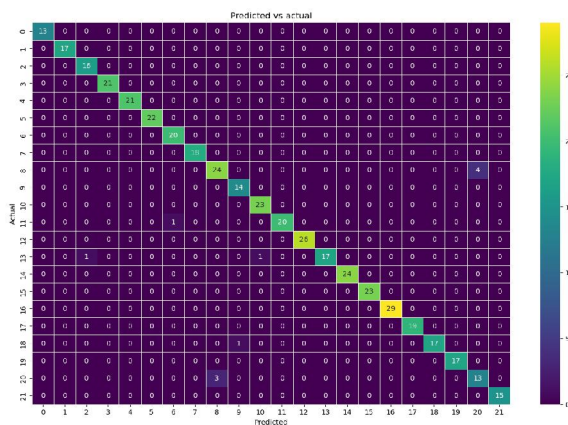

Figure 8.1: Confusion Matrix of KNN

Decision Tree
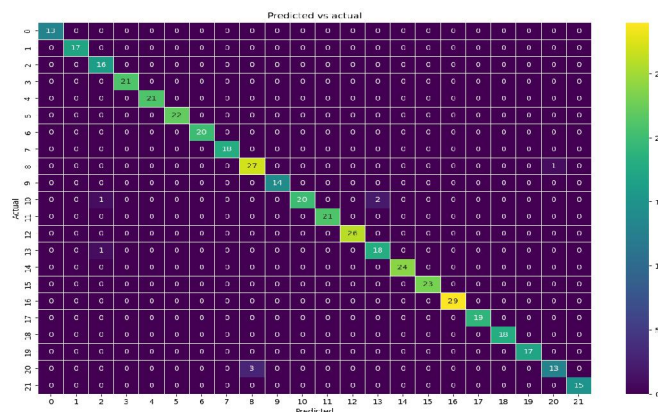

Figure 8.2: Confusion Matrix of Decision tree
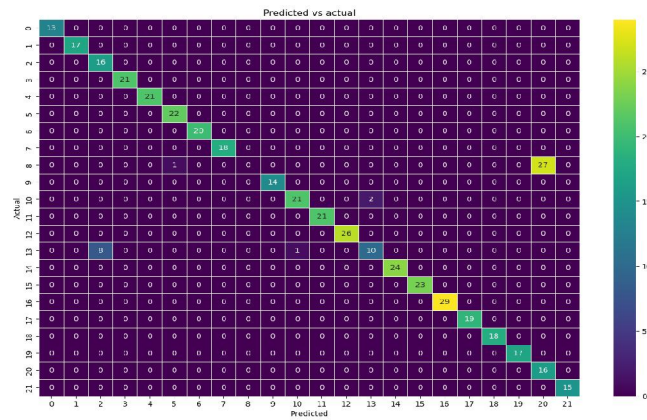
Random Forest Algorithm

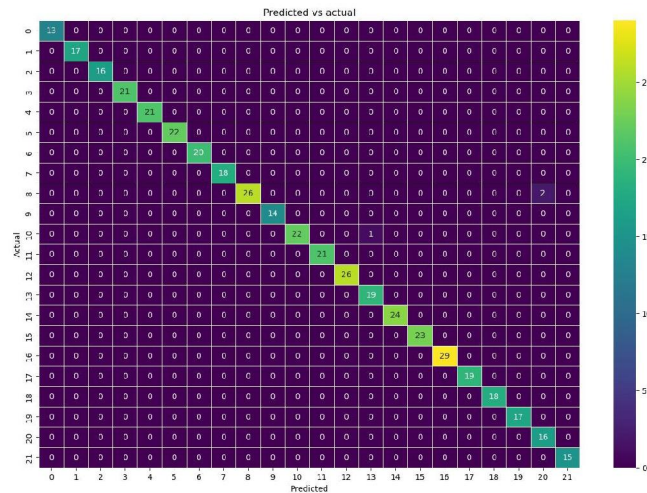Figure 8.3: Confusion Matrix of Random forest

XG Booster

Figure 8.4: Confusion Matrix of XG Booster
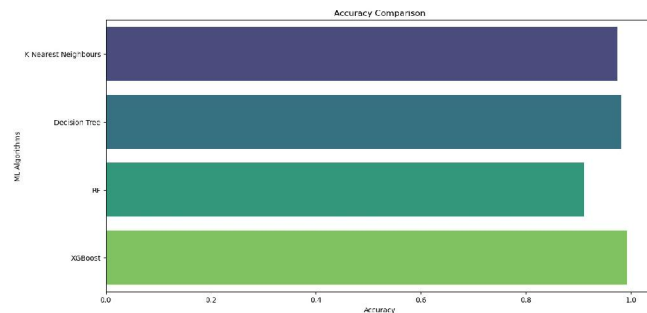
Accuracy Comparison

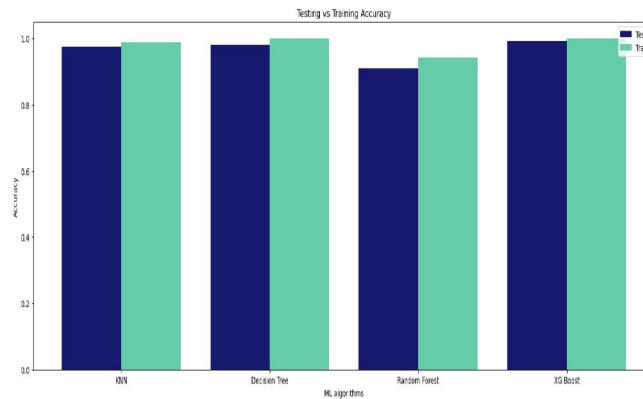Figure 8.5: Accuracy comparison

Figure 8.6 Testing vs Training Accuracy

REFERENCES

[1] Nischitha K, Dhanush Vishwakarma, Mahendra N, Ashwini, Manjuraju M.R in International Journal of Engineering Research & Technology (IJERT) , ISSN: 2278-0181, Vol. 9 Issue 08, August-2020.

[2] M.Kalimuthu, P.Vaishnavi, M.Kishore in Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020)

[3] Shilpa Mangesh Pande, Dr. Prem Kumar Ramesh in International Journal of Emerging Trends in Engineering Research, Volume 8. No. 7, July 2020

[4] Payal Gulati, Suman Kumar Jha in International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, ENCADEMS - 2020 Conference Proceedings.

[5] Ashwani Kumar Kushwaha, SwetaBhattachrya in IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555, Vol. 5, No2, April 2015]

[6] Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, B. Jabber in 2020 International Conference for Emerging Technology (INCET), Belgaum, India. Jun 5-7, 2020.

[7] SHILPA MANGESH PANDE, DR. PREM KUMAR RAMESH, ANMOL, B.R AISHWARYA, KARUNA ROHILLA, KUMAR SHAURYA in Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021).

[8] Akshata Wani1, Viraj Yadav, Ritika Todankar , Pradnyan Wade , Prof. Shikha Malik in International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056, Volume: 09 Issue: 04 | Apr 2022

[9] Nikita S.Sapike, Prof.S.S.Sambare in 2020 JETIR August 2020, Volume 7, Issue 8.

[10] Riddhi Dange, Jinisha Kande,Kunal Bhosale, Prof. Ankush Hutke in May 2022 | IJIRT | Volume 8 Issue 12 | ISSN: 2349-6002