CROP YIELD PREDICTION USING PYTHON

Kanak Sharma
Department of Computer Science,
Acropolis Institute of Technology and Research
Indore,India
kanaksharma20724@acropolis.in

Gourav Sharma
Department of Computer Science,
Acropolis Institute of Technology and Research
Indore,India
gouravsharma20702@acropolis.in

Harsh Dad
Department of Computer Science,
Acropolis Institute of Technology and Research
Indore,India
harshdad20186@acropolis.in

Harsh Jaiswal
Department of Computer Science,
Acropolis Institute of Technology and Research
Indore, India
harshjaiswal20527@acropolis.in

***Abstract*— India has witnessed adverse effects of climate change on agricultural crops over the past two decades, resulting in significant declines in their productivity. Predicting crop yields prior to harvest could offer valuable insights for policymakers and farmers, facilitating proactive measures for marketing and storage. This project aims to tackle this challenge by developing a prototype of an interactive prediction system. The system will feature a user-friendly web-based interface and employ machine learning algorithms for accurate predictions. The predicted results will be readily accessible to farmers, empowering them to make informed decisions.**

In the realm of crop prediction, various techniques and algorithms exist, with the Random Forest algorithm being selected for this project. Through the analysis of factors such as weather patterns, temperature, humidity, rainfall, and moisture levels, we aim to address the challenges encountered by farmers. In India, bolstering economic growth in agriculture is a top priority, and data mining plays a pivotal role in predicting crop yields. Data mining entails analyzing data from diverse angles and extracting crucial insights. Random Forest emerges as a popular and potent supervised machine learning algorithm, capable of handling both classification and regression tasks. It operates by constructing numerous decision trees during training and generating output based on the consensus of these trees for classification or the mean prediction for regression.

Keywords—Agriculture, Machine Learning, Crop-Prediction, Supervised Algorithms, Crop Yield, Data Mining

I. INTRODUCTION

Agriculture is the cornerstone of the Indian economy, heavily reliant on weather conditions for crop yield, with rice cultivation particularly dependent on rainfall. Providing timely forecasts for future crop productivity and conducting analyses could assist farmers in optimizing their crop production. Previously, farmers relied on past yield experiences for predictions, but now, various techniques and algorithms, such as the Random Forest algorithm, enable crop yield prediction through data analytics.

As these algorithms evolve and intersect, there's a growing array of applications and an expanding role of big data analytics in agriculture. However, despite innovative technologies, the agricultural sector faces degradation. The proliferation of hybrid and artificial products has raised concerns about health and sustainability. Moreover, modern agricultural practices often lack awareness of optimal cultivation times and locations, contributing to changes in seasonal climatic conditions that affect fundamental resources like soil, water, and air, ultimately jeopardizing food security.

Analysing these challenges, including weather, temperature, and other factors, reveals a dearth of proper solutions and technologies to address them. Nonetheless, India offers numerous avenues for bolstering economic growth in agriculture. Enhancing crop yield and quality can be achieved through various means, including data mining for predicting crop yield production.

The main objectives of this endeavour include utilizing machine learning techniques to predict crop yield, developing an easy-to-use user interface, enhancing the accuracy of crop yield prediction, and analysing

II. LITERATURE REVIEW

In [1] Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology. This paper focuses on predicting the yield of the crop based on the existing data by using Random Forest algorithm. Real data of Tamil Nadu were used for building the models and the models were tested with samples. Random Forest Algorithm can be used for accurate crop yield prediction.

In [2] Random forests for global and regional crop yield prediction. PLoS ONE Journal. Our generated outputs show that RF is an effective and adaptable machine-learning method for crop yield predictions at regional and global scales for its high accuracy and precision, ease of use, and utility in data analysis. Random Forest is the most efficient strategy and it outperforms multiple linear regression (MLR).

In [3]. Crop production Ensemble Machine Learning model for prediction. International Journal of Computer Science and Software Engineering (IJCSSE). In this paper, AdaNaive and AdaSVM are the proposed ensemble model used to project the crop production over a time period. Implementation done using AdaSVM and AdaNaive. AdaBoost increases efficiency of SVM and Naive Bayes algorithm.

In [4]. Machine learning approach for forecasting crop yield based on parameters of climate. The paper provided in International Conference on Computer Communication and Informatics (ICCCI). In the current research a software tool

named Crop Advisor has been developed as a user friendly web page for predicting the influence of climatic parameters on the crop yields.C4.5 algorithm is used to produce the most influencing climatic parameter on the crop yields of selected crops in selected districts of Madhya Pradesh. The paper is implemented using Decision Tree.

In[5]. Prediction On Crop Cultivation. International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) Volume 5, Issue 10, October 2016. Presently, soil analysis and interpretation of soil test results is paper based. This in one way or another has contributed to poor interpretation of soil test results which has resulted into poor recommendation of crops, soil amendments and fertilizers to farmers thus leading to poor crop yields, micro-nutrient deficiencies in soil and excessive or less application of fertilizers. Formulae to Match Crops with Soil, Fertilizer Recommendation.

In [6]. Analysis of Crop Yield Prediction by making Use Data Mining Methods. IJRET: The paper provided in International Journal of Research in Engineering and Technology. In this paper the main aim is to create a user-friendly interface for farmers, which gives the analysis of rice production based on the available data. For maximizing the crop productivity various Data mining techniques were used to predict the crop yield. Such as K-Means algorithm to forecast the pollution factor in the atmosphere.

In [7]. Applications of Machine Learning Techniques in Agricultural Crop Production. Indian Journal of Science and Technology, Vol 9(38), DOI:10.17485/ijst/2016/v9i38/95032, October 2016. From GPS based colour images is provided as an intensified indistinct cluster analysis for classifying plants, soil and residue regions of interest. The paper includes various parameters which can help the crop yield for better enhancement and ratio of the yield can be increased during cultivation.

In [8] In this paper, we present a comprehensive review of research dedicated to the application of machine learning in agricultural production systems. Machine learning (ML) has emerged together with big data technologies, techniques, methods and high-performance computing to generate new opportunities to unravel, quantify, and analyse data intensive processes in agricultural operational sectors. By using Support Vector Machines (SVP) the Paper is Implemented.

In [9]. A Study to Determine Yield for Crop Insurance using Precision Agriculture on an Aerial Platform. Symbiosis Institute of Geoinformatics Symbiosis International University 5th & 6th Floor, Artur Centre, Gokhale Cross Road, Model Colony, Pune – 411016. Precision agriculture (PA) is the application of geospatial methodologies and remote sensors to identify variations in the field and to deal with them using different strategies. The causes of variability of crop growth in an agricultural field might be due to crop stress, irrigation practices, incidence of pest and disease etc. The Paper is Implemented using Ensemble Learning (EL).

In [10]. Random Forests for Global and Regional Crop Yield Predictions. institute on the Environment, University of Minnesota, St. Paul, MN 55108, United States of America. The generated outputs show that RF is an effective and different machine-learning method for crop yield predictions at regional and global scales for its high accuracy. The Paper is Implemented using k-nearest neighbour, Support Vector Regression (SVG).

III. METHODOLOGY

Data serves as a crucial component in any Machine Learning System. In our implementation, we chose to focus on Maharashtra State in India due to its diverse climate. As climate conditions vary from place to place, it was essential to gather data at the district level. Historical data concerning both crops and climate in specific regions was necessary for our system. We collected crop data for each district of Maharashtra from www.data.gov.in and climate data from www.imd.gov.in. The climatic parameters that significantly impact crops include precipitation, temperature, cloud cover, vapor pressure, and wet day frequency. Therefore, we gathered monthly data on these climatic parameters.

Dataset Collection: In this phase, we gathered data from various sources and prepared datasets for analysis. The provided dataset was utilized for analytics purposes, including descriptive and diagnostic analyses. Online abstract sources such as Data.gov.in and indiastat.org provided yearly abstracts of crops for at least ten years, typically exhibiting the behaviour of random time series. We combined primary and necessary abstracts to create comprehensive datasets.

Random Forests for Global and Regional Crop Yield Predictions: Random Forests, a machine learning algorithm, were employed for global and regional crop yield predictions.

Data Partitioning: The entire dataset was partitioned into two parts: 75% of the dataset was used to train the model, while the remaining 25% was reserved for testing the model, ensuring its accuracy in predicting future events. Table Type Styles

IV. IMPLEMENTATION

Python: 3.x (we used Python 3.8.8 in this project)PyCharm: PyCharm is an integrated development environment (IDE)for Python programming. It is developed by JetBrains, a company known for creating popular development tools for various programming languages. PyCharm is specifically designed to enhance the productivity of Python developers by providing a rich set of features and tools for Python development.

Data Availability: Assess the availability and accessibility of necessary data sources, such as historical agricultural data, weather data, and soil data. Ensure that data quality and reliability meet project requirements.

Streamlit: Streamlit lets you turn data scripts into shareable web apps in minutes, not weeks. It's all Python, open-source, and free!

Sklearn: Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modelling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. Scikit-learn is probably the most useful library for machine learning in Python. The sklearn library contains a lot of efficient tools for machine learning and statistical modelling including classification, regression, clustering and dimensionality reduction.

Crops Production Data: he area of study has fertile soil because the district is surrounded by the volcanic mountains, and this makes the region favourable for agriculture. Maize and Irish potatoes are the most grown crops in this region. Due to the importance of the two crops to the economy of the country and food security, the district decided to consolidate the agriculture land under the CIP. This strategy has the target of increasing the production of two crops from 5.8 tonnes ha−1 and 29.53 tones ha−1 in the agriculture year 2017/18 to10 tones ha−1 and 42 tones ha−1 in 2022/23 for maize and Irish potatoes, respectively. The agriculture of the two crops is carried out in two seasons: the first season (A) that starts in September and end in January, whereas the second season start in February to June. The two crops are grown alternatively. Irish potatoes are harvested after 3–4 months after planting whereas maize is harvested after 5–6 months.

The data for the crop yields were gathered from NISR and from different cooperatives of farmers in the area of study. The collected data were from agriculture year 2005/2006 to 2020/2021.(Fig:1)

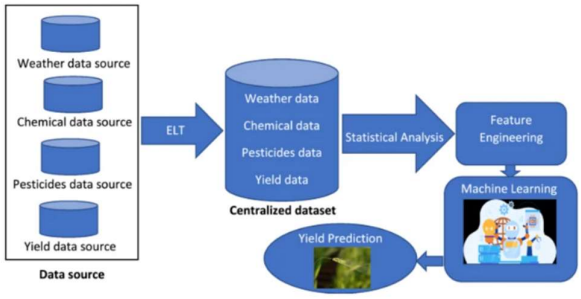

Fig 1: Crop yield prediction pipeline overview

Weather Data: The key weather parameters that have an impact on the crop's development are precipitation, air temperature, air humidity, solar radiation, and wind speed [11]. However, due to the unavailability of the data from the sources for the entire period of the study, this study used rainfall and air temperature since these were the ones available for the period considered in this study.

Rainfall: Precipitation data were collected from Rwanda meteorology agency. Since crops have various water needs according to their stage of development, the monthly total rainfall have been calculated from the daily precipitation to identify the correlation between monthly water requirements and crop production.

Air Temperature: In this study, the monthly mean temperatures have been used since the temperature requirement is different according to the crop stage. The following tables (Table 1 and Table 2) summarize the datasets for Irish potato and maize, respectively.

Random Forest: Random Forest is one of the prominent supervised machine learning algorithms. It can work on both categorical and regression problems with the use of numerous trees, the bootstrap method, and aggregation. The term "decision tree" comes from the way they appear to flow similar to trees. They begin at the tree's base and proceed through splits with uncertain results until they reach a leaf node, where the outcome is revealed [12].

V. CONCLUSION AND FUTURE SCOPE

Based on the climatic input parameters the present study provided the demonstration of the potential use of data mining techniques in predicting the crop yield based. The developed webpage is user friendly and the accuracy of predictions are above 75 per cent in all the crops and districts selected in the study indicating higher accuracy of prediction. By providing climatic data of that place the user-friendly web page developed for predicting crop yield can be used by any user their choice of crop.

VI. ACKNOWELDGEMNT

We extend our heartfelt gratitude to all those who contributed to the completion of this research paper. Special thanks to Professor Krupi Saraf,, whose expertise and guidance have been invaluable in shaping this paper. Lastly, we thankful to our family, friends, and colleagues for their unwavering support and encouragement.

REFERENCES

[1] P.Priya, U.Muthaiah M.Balamurugan.Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology.

[2] J.Jeong, J.Resop, N.Mueller and team.Random forests for global and regional crop yield prediction.PLoS ONE Journal.

[3] Narayanan Balkrishnan and Dr. Govindarajan Muthukumarasamy.Crop production Ensemble Machine Learning model for prediction. International Journal of Computer Science and Software Engineering (IJCSSE).

[4] S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh.Machine learning approach for forecasting crop yield based on climatic parameters. International Conference on Computer Communication and Informatics (ICCCI).

[5] Shweta K Shahane , Prajakta V Tawale.Prediction On Crop Cultivation. IInternational Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) Volume 5, Issue 10, October 2016

[6] D Ramesh ,B Vishnu Vardhan. Analysis Of Crop Yield Prediction Using Data Mining Techniques. IJRET: International Journal of Research in Engineering and Technology

[7] Subhadra Mishra,Debahuti Mishra, Gour Hari Santra. Applications of Machine Learning Techniques in Agricultural Crop Production. Indian Journal of Science and Technology, Vol 9(38), DOI:10.17485/ijst/2016/v9i38/95032, October 2016

[8] Konstantinos G. Liakos,Patrizia Busato,Dimitrios Moshou, Simon Pearson ID,Dionysis Bochtis. Machine Learning in Agriculture. Lincoln Institute for Agri-food Technology (LIAT), University of Lincoln, Brayford Way, Brayford Pool,Lincoln LN6 7TS, UK, spearson@lincoln.ac.uk.

[9] Baisali Ghosh. A Study to Determine Yield for Crop Insurance using Precision Agriculture on an Aerial Platform. Symbiosis Institute of Geoinformatics Symbiosis International University 5th & 6th Floor, Atur Centre, Gokhale Cross Road, Model Colony, Pune – 411016.

[10] Jig Han Jeong, Jonathan P. Resop, Nathaniel D. Mueller, David H. Fleisher ,Kyungdahm Yun, Ethan E. Butler,Soo-Hyung Kim. Random Forests for Global and Regional Crop Yield Predictions. Institute on the Environment, University of Minnesota, St. Paul, MN 55108, United States of America.

[11] Chakraborty, D.; Saha, S.; Sethy, B.K.; Singh, H.D.; Singh, N.; Sharma, R.; Chanu, A.N.; Walling, I.; Anal, P.R.; Chowdhury, S.; et al. Usability of the Weather Forecast for Tackling Climatic Variability and Its Effect on Maize Crop Yield in Northeastern Hill Region of India. Agronomy 2022, 12, 2529.

[12] Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32. [Google Scholar] [CrossRef] [Green Version]