International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 6 ǁ June 2023 ǁ PP. 507-512

Cereal Production Analysis Using ML

Shalini K[1], Davan Devang H S[2], Gagan C[3], Kruthik B[4], Naresh Kumar P[5]
1 - Assistant Professor, Dept of CSE, Dr Ambedkar Institute of Technology, Bangalore-560056 2, 3, 4, 5 - Under Graduates, Dept of CSE Dr Ambedkar Institute of Technology, Bangalore-560056

Abstract
Many farmers are intrigued in gaining an estimate of the cost of the crop after the produced cereal has been reaped. The core of theIndian economy is agriculture. The idea that "traditional farming in the regular months has been distorted and crops have been ruined" is one that is often heard today. In addition to incurring economic losses, this is the main reason of farmer suicide. Now thatagriculture demands guidance, it is time for technology to lead the evolution. Favorable soil conditions, ambient rainfall, and temperature are required for a crop to grow. As a result, it is currently difficult for farmers and the general public to forecast monthsfor planting and agricultural output due to abnormalities in temperature and rainfall as a result of climate change, for instance, rainsin December and January or erratic temperatures. As a result, we have developed an analysis by crop prediction that takes into account a number of variables. Machine learning has demonstrated that the data we have is essential for making predictions. Agriculture is a subject where machine learning is expanding and has a bright future.

Keywords: *Machine Learning (ML), Prediction, Evaluation, Random Forest, Decision Trees, Feature Selection*

Date of Submission: 13-06-2023 Date of acceptance: 28-06-2023

I. INTRODUCTION

The agriculture sector has been revolutionized by cereal production analysis using machine learning (ML), which has given farmers effective tools to increase crop yields and make wise decisions. A expanding population depends on cereal crops like wheat, rice, and maize, yet maximizing their production can be challenging. The application of ML approaches to agricultural research opens up a brand-new field, utilizing the power of data analysis to generate insightful discoveries and promote environmentally friendly farming methods. Farmers can use historical data on a variety of variables that affect cereal yield with ML algorithms. These variables include pest infestations, soil quality, weather, and fertilizer use. Farmers can make data-driven decisions by feeding this data into ML models to find patterns and linkages. For instance, ML systems may examine enormous datasets to recommend accurate fertilizer application rates, warn farmers of impending disease outbreaks, and forecast the best periods for planting. ML enables farmers to adapt their farming practices and take proactive actions by supplying accurate and timely information. This, in turn, improves yields and reduces resource waste. This can aid farmers in conducting more efficient agricultural analyses and production. In conclusion, ML-based cereal production analysis is a revolutionary method that gives farmers the resources they need to maximize crop yields and advance sustainable agriculture. Farmers may make better decisions with ML thanks to historicaldata and sophisticated algorithms, which boost productivity and resource management. In a world with changing environmental concerns, combining technology and agriculture has enormous promise for guaranteeing food security and sustainability.

II. RELATED WORKS

"Crop yield prediction using machine learning: A systematic literature review" by Paudel et al. (2020) - This study provides a comprehensive review of machine learning techniques applied to crop yield prediction, highlighting their strengths, limitations, andfuture research directions.

"Spatio-temporal analysis of crop yield using machine learning techniques" by Luo et al. (2019) - The research explores the use of machine learning methods to analyze the spatio-temporal patterns of crop yield, enabling better understanding and prediction of cereal production at various locations and time periods.

"Machine learning-based crop yield prediction: Methods, applications, and future perspectives" by Dall'Asta et al. (2021) - This work presents an overview of machine learning-based methods for crop yield prediction, discussing their applications, challenges,and potential future advancements in the field.

“Crop yield estimation using machine learning: A review" by Shah et al. (2019) - The study reviews the application of machine learning techniques in crop yield estimation, discussing the key factors influencing yield and the role of ML models in improving prediction accuracy.

"Remote sensing-based crop yield prediction: A review" by Wang et al. (2021) - This review paper focuses on the use of remote sensing data combined with machine learning techniques for crop yield prediction, highlighting the advancements and challenges in this interdisciplinary field.

1. Title: "Machine Learning Techniques for Crop Yield Prediction: A Comprehensive Review" Authors: Liu, S., et al. Published in: Computers and Electronics in Agriculture, 2021 Summary Liu et al. (2021) conducted a comprehensive review of machine learning techniques applied to crop yield prediction, including cereal crops. They discussed various methodologies, such as regression models, ensemble methods, and deep learning, along with the data sources used. The paper highlighted the challenges faced in utilizing machine learning for cereal production analysis, including data quality,model interpretability, and scalability. Overall, the review provided insights into the potential of machine learning for improving cereal production analysis and emphasized the need for further research to overcome the existing limitations.
2. Title: "Applications of Machine Learning in Agriculture and Allied Fields: A Review" Authors: Kumar, S., et al. Published in: Information Processing in Agriculture, 2020 Summary Kumar et al. (2020) presented an extensive review on the applications of machine learning in agriculture, which also covered cereal production analysis. The paper discussed various machine learning algorithms employed for predicting crop yield, such as support vector machines, random forests, and neural networks. They explored different data sources used in agricultural analysis, including satellite imagery, weather data, and soil information. The review highlighted the performance of machine learning models in predicting cereal crop yields and provided valuable insights into the challenges and opportunities associated with employing these techniques incereal production analysis.
3. Title: "Review on Machine Learning Applications in Agriculture and Food Supply Chain: Datasets, Challenges, and FuturePerspectives" Authors: Muzammal, M., et al. Published in: Computers and Electronics in Agriculture, 2020 Summary: Muzammal et al. (2020) conducted a review on machine learning applications in agriculture and the food supply chain, including cereal production analysis. The paper discussed the datasets used in machine learning-based agricultural analysis, such as remote sensing data, farm management systems, and market data. It identified challenges related to data availability,data quality, and algorithm selection. The review provided insights into the potential future perspectives of machine learning in cereal production analysis, including the integration of Internet of Things (IoT) technologies, big data analytics,and precision agriculture. It emphasized the need for addressing challenges and implementing innovative solutions to enhance cereal production efficiency.
4. "Machine Learning Techniques for Crop Disease Detection and Diagnosis: A Review" Authors: Raza, S. E. A., et al. Published in: Computers and Electronics in Agriculture, 2019 Summary Raza et al. (2019) focused on machine learning techniques for crop disease detection and diagnosis. Although not specific to cereal production analysis, the insights fromthis review paper are relevant, as diseases significantly impact cereal crops. The paper discussed various machine learningapproaches, including image-based disease detection, sensor-based monitoring, and expert systems. It highlighted the importance of accurate and timely disease diagnosis in managing crop diseases effectively. The review emphasized the potential of machine learning in aiding disease detection and decision-making processes in cereal production, leading to improved crop health and yield.
5. Title: "Machine Learning for Crop Yield Prediction: A Review" Authors: Banerjee, S., et al. Published in: Computers and Electronics in Agriculture, 2019 Summary Banerjee et al. (2019) conducted a review specifically on machine learning for crop yield prediction, including cereals. The paper discussed different machine learning models, such as regression algorithms, decision trees, and neural networks, used for predicting cereal crop yields. It provided an overview of data sources utilized, including weather data, satellite imagery, and historical yield records. It highlighted the promising performance of machine learning models in predicting cereal crop yields and emphasized the need for further research in this domain to improve accuracy and robustness.

III. EXPIREMENT

To compare the behaviours of Features, the experiment conducted was focused on the evaluation of the algorithms.

Project Setup

First, we need to set up the project directory and create a virtual environment to manage our dependencies. We can create a new directory called "Cereal Production Analysis" and navigate to it in the terminal. Then, we can create a virtual environment using thecommand "python-m venv venv". Next, we activate the virtual environment by running "source venv/bin/activate". Now, we can install Flask and other required libraries using pip. Where the flask is confined to only that particular folder. If we want to create other folder we have to again create the project's dedicated folder.

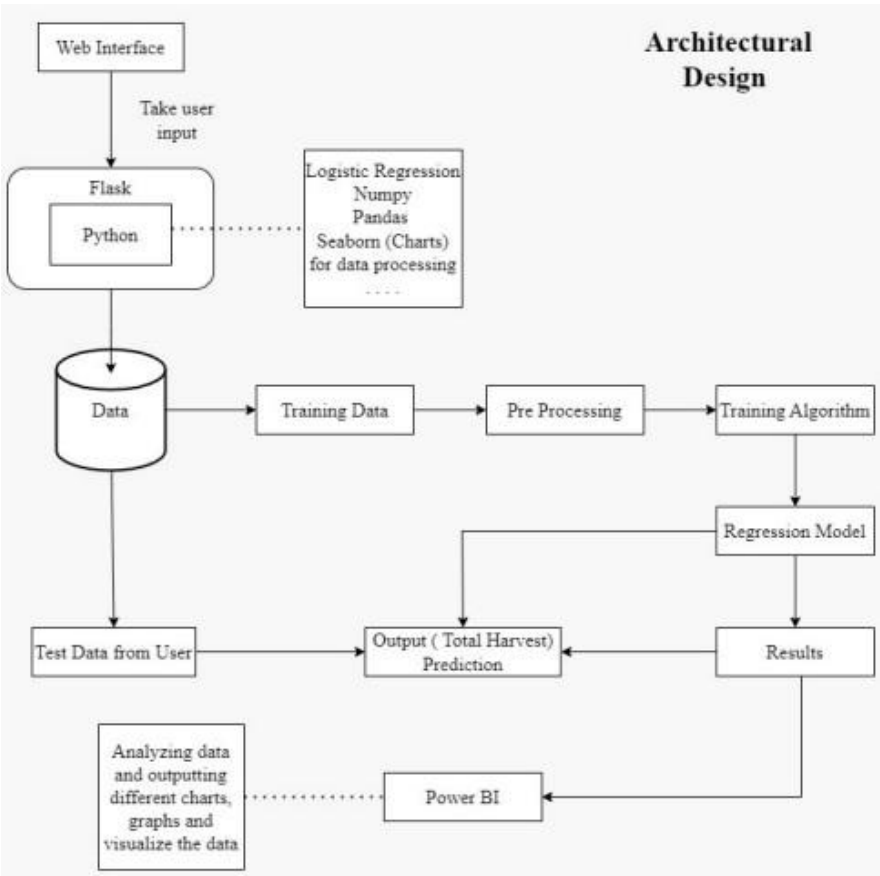
Dataset

We need to collect relevant data on cereal production, such as temperature, rainfall, soil composition, and crop yield. Let's say we have a dataset with columns like plot size, varieties, seed in Kgs, urea, phosphate, pesticide, labour. We need to clean the data by removing any missing values or outliers. We can also transform and format the data to ensure consistency and compatibility with ML algorithms, such as converting categorical variables into numerical representations.

The data set is collected from Kaggle. The features mentioned are:

Total_working_labour_hrs, phosphate_kg, price_phosphate, seed_kg, urea_kg, wage_hrs.
Feature Selection (Is done through distplot and mutual information through Pandas) Selection of Features for Prediction (Here Co- linearity check is done to select features for prediction of total HARVEST as well as PRICE)

IV. ARCHITECTURE DIAGRAM



The initial step in the cereal production analysis using ML involves collecting user input through a web interface. Flask, a Python framework, is utilized as the back-end framework for developing the application. Flask is a lightweight and modular framework that simplifies the development process by providing pre-built modules and libraries, eliminating the need for low-level code implementation, such as handling protocols and managing threads. Python, with its extensive ecosystem, offers powerful tools for data processing. Logistic regression is employed as the ML algorithm to train the regression model. NumPy and Pandas libraries are utilized for efficient data manipulation and analysis. Seaborn, a data visualization library, aids in creating informative and visually appealing plots, enhancing the analysis and interpretation of data.

V. IMPLEMENTATION

1. Project Setup: First, we need to set up the project directory and create a virtual environment to manage our dependencies. We can create a new directory called "Cereal Production Analysis" and navigate to it in the terminal. Then, we can create a virtual environment using the command "python -m venv ". Next, we activate the virtual environment by running "sourcevenv/bin/activate". Now, we can install Flask and other required libraries using pip. Where the flask is confined to only that particular folder. If we want to create other folders, we have to again create the project's dedicated folder.
2. Data Collection and Preprocessing: We need to collect relevant data on cereal production, such as temperature, rainfall, soil composition, and crop yield. Let's say we have a dataset with columns like plot size, varieties, seed in Kgs, urea, phosphate, pesticide, labor. We need to clean the data by removing any missing values or outliers. We can also transform and format the data to ensure consistency and compatibility with ML algorithms, such as converting categorical variables into numerical representations. Available features (Here all the features are listed out from the dataset)
3. ML Model Selection: We need to choose suitable ML algorithms for our cereal production analysis. Let's say we decide touse a KNeigborsClassifier. We can import the necessary libraries, create an instance of the KNeigborsClassifier and set any desired hyperparameters.
4. Model Training and Evaluation: Now, we split our dataset into training and testing sets. We use the training set to train our random forest regressor by fitting it to the input features and the target variable (crop harvest). We adjust the hyperparameters to optimize the model's performance. Once trained, we evaluate the model's performance using metrics such as mean squared error or R-squared value. If the R-squared value is closer to 1 then the model is chosen for predictionor not.
5. Flask App Structure: Let's set up the structure of our Flask application. We create directories like "static" for storing staticfiles (CSS, JavaScript) and "templates" for HTML templates. These directories will help organize our files and make it easier to serve our front-end files.
6. Front-End Development: Now, we can design and develop the front-end user interface using HTML, CSS, and JavaScript within the Flask templates. We create forms and input fields to allow users to enter data. We can also add buttons to submit the data for analysis and display the results. Front end development (Where a form is created to take input from the user and predict the output through a list of input parameters provided to the prediction model)
7. Flask Routes: We define Flask routes to handle incoming requests from the front-end. For example, we create a route called "/predict" that handles a POST request with the user input data. Inside this route, we preprocess the input data, pass it to the trained Linear Regression Model, and obtain the predicted crop yield. We then return this result to the front-end for display.
8. Visualization and Output: We present the results of the cereal production analysis to the user through the front-end interface. For example, we can dynamically update a section of the webpage to display the predicted crop yield based on the user's input. Through return statement the output of the prediction can be displayed in the HTML file through routing and rendering of the page.
9. Testing and Deployment: It's important to thoroughly test our application to ensure it functions correctly. We can simulate different scenarios by entering different values for temperature, rainfall, and soil pH and checking if the predicted yield aligns with our expectations. Once we are satisfied with the testing, we can deploy the Flask application to a hosting serveror cloud platform, making it accessible online. If the project is up and running on the localhost then the project can be deployed on GAE for global usage and generating a link through which we can access the site from wherever we want.

VI. RESULTS

Testing and Deployment: It's important to thoroughly test our application to ensure it functions correctly. We can simulate differentscenarios by entering different values for temperature, rainfall, and soil pH and checking if the predicted yield aligns with our expectations. Once we are satisfied with the testing, we can deploy the Flask application to a hosting server or cloud platform, making it accessible online. If the project is up and running on the localhost then the project can be deployed on GAE for global usage and generating a link through which we can access the site from wherever we want.

Screenshots:

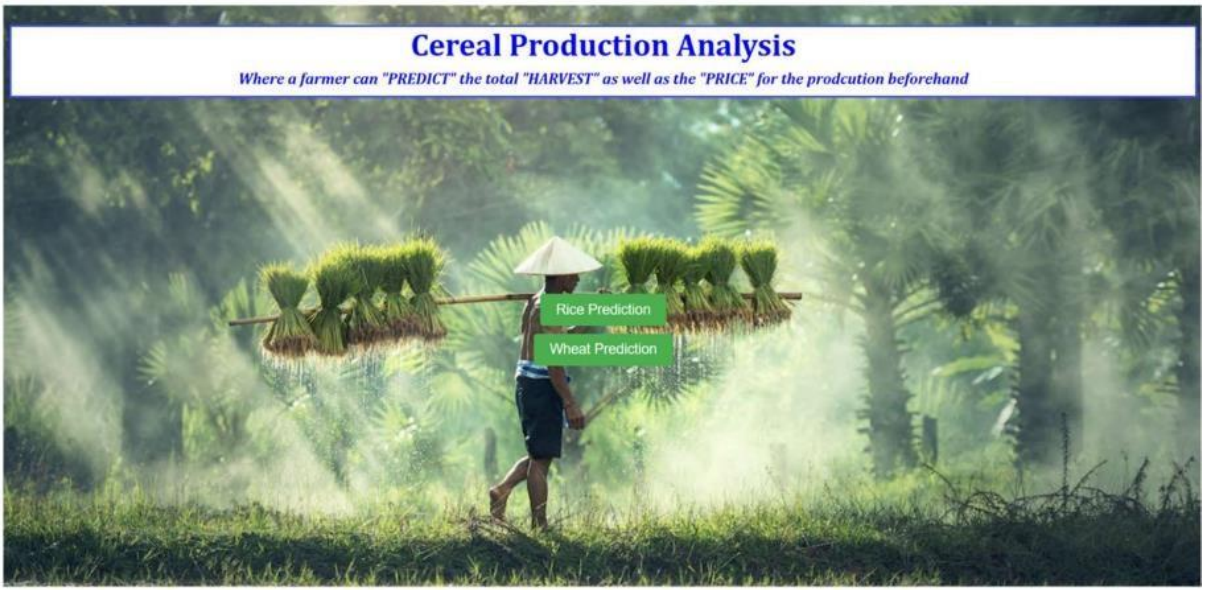


Fig 1: Home page

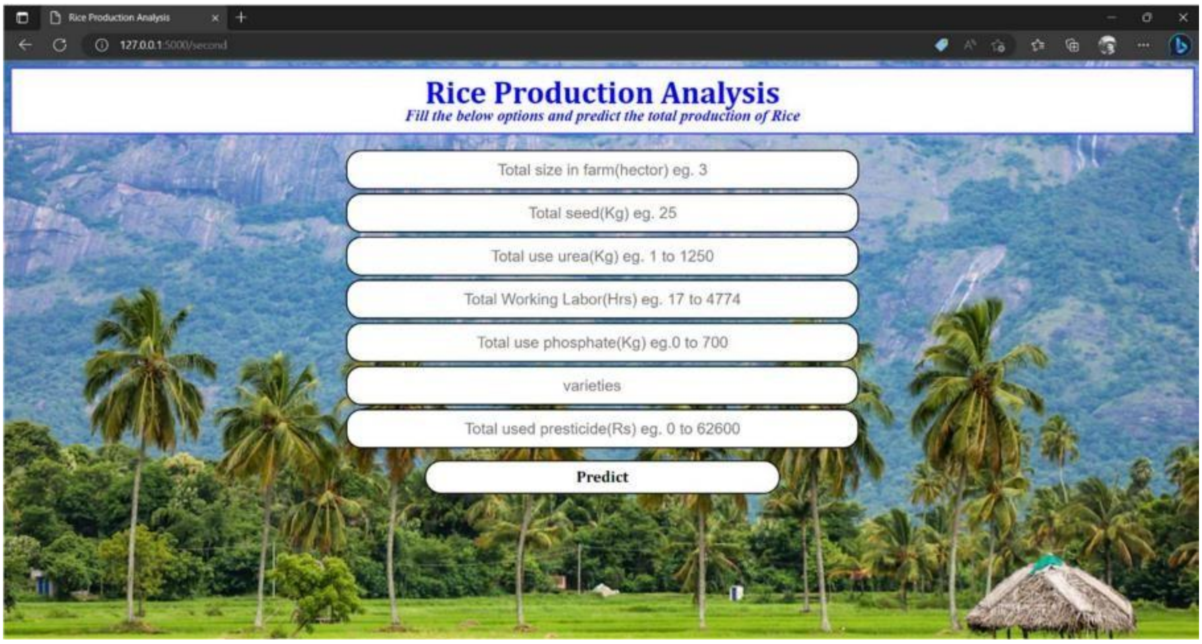


Fig 2: Rice Production Analysis Page

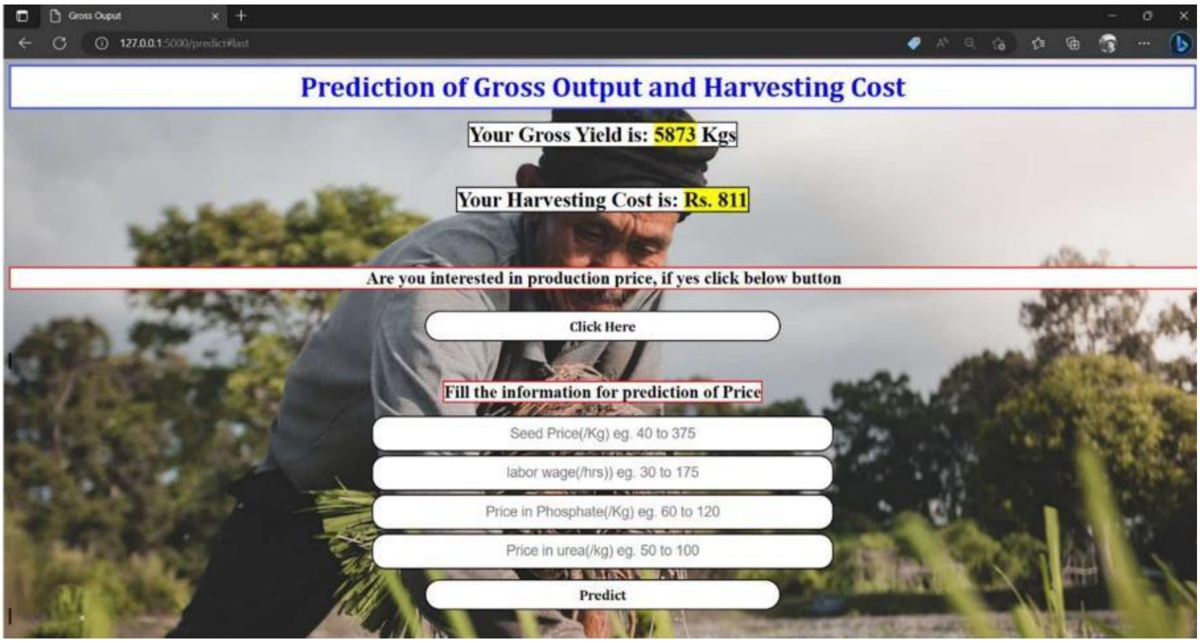


Fig 3: Prompt for Price/Kg Prediction of Rice

VII. CONCLUSION

Cutting-edge machine learning techniques are used in this study to completely transform crop forecasting and yield estimation. Accurate crop predictions for certain districts can be made using the Random Forest classifier. The method effectively examines past data, giving farmers enlightening information to select crops intelligently. This innovative strategy improves agricultural practises, resulting in efficient and productive harvests. Increased agricultural productivity, promoting economic growth within the agricultural sector, and maximising crop yield rates are potential effects of this effort. The technology also provides farmers with crucial information regarding the best crops to grow in their particular soil conditions, enabling them to modify their farming practises. This project seeks to increase total agricultural output by identifying the most promising crops for each district while optimising resource allocation and reducing hazards. This programme promotes overall food security while also supporting sustainable farming methods and regional economic growth.

REFERENCES

[1]. Research Paper: "Crop yield prediction using machine learning: A systematic literature review" by S. Singhetal. (Link: [ResearchPaper](https://www.sciencedirect.com/science/article/pii/S235234091931271X)).

[2]. Research Paper: "Machine Learning Approaches for Crop Yield Prediction: An Empirical Study" by R. Kaur et al. (Link: [Research Paper] (https://ieeexplore.ieee.org/abstract/document/8665439)).

[3]. Website: Kaggle-Cereal Production Dataset (Link:[Kaggle Dataset] (https://www.kaggle.com/dorbicycle/world-foodfeed-production)).

[4]. Research Paper: "Crop Yield Prediction Based on Machine Learning Algorithms" by X. Chen et al. (Link: [Research Paper] (https://ieeexplore.ieee.org/abstract/document/8992444))

[5]. Website: UCI Machine Learning Repository- Crop Production Dataset (Link: [UCIDataset] (https://archive.ics.uci.edu/ml/datasets/crop+production).

[6]. Research Paper: "Machine Learning Techniques for Crop Yield Prediction: A Comparative Study" by A. Jatav et al. (Link: [Research Paper] (https://link.springer.com/chapter/10.1007/978-981-15-0684-7_26)).

[7]. Research Paper: "Crop Yield Prediction Model Using Machine Learning Algorithms" by A. Choudhuryetal. (Link: [ResearchPaper](https://www.sciencedirect.com/science/article/pii/S1877050919314933)).