International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

INDIAN CROP YIELD ANALYSIS AND PREDICTION

[1] K Vijay, [2] B Sai Vamshi, [3] S Ramya Sri , [4] K Naveen
[1]AssistantProfessor, [234]Students
Department of Data Science
Siddhartha Institute of Technology & Sciences, Narapally
vijaykoraveni.cse@siddhartha.co.in, 23TQ1A6714@siddhartha.co.in, 23TQ1A6721@siddhartha.co.in, 23TQ1A6734@siddhartha.co.in,

ABSTRACT

Agriculture is one of the most important sectors of the Indian economy and provides employment and food resources for a large population. Crop yield can vary significantly due to factors such as rainfall, temperature, soil characteristics, irrigation, crop type, fertilizer usage, and cultivation area. Analyzing these factors can help identify patterns in agricultural production and support better planning.

The Indian Crop Yield Analysis and Prediction system is designed to analyze historical agricultural data and predict crop yield using data analytics and Machine Learning techniques. The system uses information such as state, district, crop type, cultivated area, production, rainfall, temperature, soil properties, irrigation, and fertilizer usage. Data preprocessing is performed to clean and prepare the dataset for analysis.

The system performs exploratory data analysis to understand crop production and yield patterns across different states, districts, crops, and years. Statistical measures and visualizations are used to compare crop performance and identify changes in agricultural productivity. Charts, graphs, maps, and dashboards can provide an easy-to-understand representation of the collected agricultural information.

The prediction module uses selected agricultural features to train a Machine Learning model for estimating crop yield. Different algorithms can be evaluated based on suitable performance metrics, and the model with appropriate performance can be used for prediction. The system can accept relevant agricultural inputs and generate an estimated yield value for the selected crop or region.

Overall, the proposed system combines agricultural data analysis with predictive modeling to provide useful insights into Indian crop production. It can support farmers, agricultural researchers, analysts, and policymakers in understanding historical yield patterns and planning agricultural activities. Future enhancements can include real-time weather data, satellite imagery, soil sensors, crop disease detection, irrigation recommendations, and advanced yield forecasting.

I. INTRODUCTION

Agriculture plays a major role in India's economy and is essential for maintaining food security and supporting rural communities. India has diverse agricultural regions with different soil conditions, climates, rainfall patterns, and crop-growing practices. As a result, crop productivity can vary significantly between regions and seasons.

Crop yield is influenced by several factors, including rainfall, temperature, soil fertility, irrigation, cultivated area, fertilizer application, crop variety, and farming practices. Changes in weather conditions can also affect agricultural production. Understanding the relationship between these factors and crop yield can help improve agricultural planning.

Traditional crop-yield analysis often relies on historical records, surveys, and manual calculations. Although these methods provide valuable information, analyzing large agricultural datasets manually can be time-consuming. It can also be difficult to identify complex relationships between multiple environmental and agricultural variables.

Data analytics and Machine Learning provide techniques for analyzing large datasets and identifying patterns that may not be easily visible through traditional methods. Historical crop data can be used to study production trends, regional differences, crop performance, and relationships between environmental factors and yield.

The proposed Indian Crop Yield Analysis and Prediction system integrates data preprocessing, exploratory analysis, visualization, and Machine Learning-based prediction. It provides users with insights into crop production and yield patterns across Indian regions and generates estimated crop yields based on available agricultural features. The system can serve as a decision-support tool for agricultural analysis and planning.

II. LITERATURE SURVEY

Crop yield prediction has been widely studied using statistical analysis, agricultural datasets, remote sensing, and Machine Learning techniques. Historical agricultural datasets generally contain information about crop type, cultivated area, production, location, year, rainfall, temperature, soil properties, and other environmental factors. Researchers use these datasets to understand factors associated with agricultural productivity.

Traditional statistical approaches have been used to analyze relationships between crop yield and variables such as rainfall, temperature, cultivated area, and fertilizer usage. Regression techniques can estimate yield based on selected agricultural factors. These methods provide interpretable relationships and can be useful when the data follows relatively stable patterns.

Weather conditions are important variables in agricultural yield analysis. Rainfall, temperature, humidity, and other climatic conditions can influence crop growth and production. Integrating weather information with historical crop records can provide additional information for understanding seasonal and regional yield variations.

Geographic and regional analysis is another important area of agricultural research. Crop productivity can vary between states and districts because of differences in soil, climate, irrigation availability, farming practices, and crop selection. Visualization techniques such as maps and regional charts can help identify areas with different production and yield patterns.

Machine Learning techniques have increasingly been applied to crop-yield prediction. Algorithms such as Linear Regression, Decision Trees, Random Forest, Gradient Boosting, and other models can learn relationships between historical agricultural features and crop yield. Model performance can be evaluated using metrics such as MAE, MSE, RMSE, and R^2 depending on the prediction task.

Recent agricultural systems combine Machine Learning with weather data, satellite imagery, soil information, and sensor data to improve crop prediction. Such systems can support precision agriculture and resource planning. The proposed system follows this direction by combining historical Indian agricultural data, exploratory analysis, visualization, and Machine Learning-based crop-yield prediction.

III. SYSTEM ANALYSIS

The Indian Crop Yield Analysis and Prediction system is designed to process historical agricultural data and generate useful crop-yield insights. The system accepts information such as state, district, crop type, cultivated area, production, rainfall, temperature, soil properties, irrigation, and fertilizer usage. The input dataset is first

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

validated to ensure that required fields and supported formats are available. A preprocessing module handles missing values, duplicate records, inconsistent units, and invalid entries. Agricultural dates and locations are standardized to support regional and time-based analysis. Feature engineering is performed to create useful attributes for analytical and predictive modeling. The analytics engine calculates crop production, yield values, growth trends, and regional comparisons. Exploratory analysis identifies relationships between crop yield and agricultural or environmental factors. Visualization components present results through charts, graphs, maps, tables, and KPI cards. A Machine Learning model is trained using historical crop data and selected features. The prediction module estimates crop yield for new agricultural inputs based on the trained model. Model performance is evaluated using suitable regression metrics before the prediction model is used. The final system presents analytical findings and predicted yield values through an interactive dashboard.

Existing System

Existing agricultural information systems mainly focus on collecting crop production statistics, maintaining agricultural records, and providing basic reports. Government departments, research institutions, and agricultural organizations may maintain historical datasets containing crop production and area information. These systems are useful for record keeping but may provide limited interactive analysis and predictive capabilities. Farmers and analysts may need to use separate tools to examine relationships between weather, soil, irrigation, and crop yield. Manual analysis of large agricultural datasets can require significant time and effort. Traditional reports may not clearly represent regional and seasonal variations in crop productivity. Historical production information may be available without integrated Machine Learning-based yield prediction. Comparing crops across states, districts, and years can also require additional processing. Weather and soil information may be stored separately from crop production data. Data visualization may be limited to static charts or tables. As a result, users may find it difficult to obtain comprehensive insights from multiple agricultural factors. An integrated analytics and prediction platform can provide a more efficient approach to crop-yield analysis.

Disadvantages of Existing System

- Limited crop-yield prediction capabilities.
- Manual analysis of agricultural datasets.
- Difficult to process large datasets efficiently.
- Limited interactive visualization.
- Weather and crop data may be stored separately.
- Difficult to identify regional yield patterns.
- Limited analysis of soil-related factors.
- Static reports may provide insufficient insights.
- Time-consuming comparison of crops and regions.
- Limited support for predictive decision-making.
- Difficult to identify complex relationships between variables.
- Limited integration of multiple agricultural data sources.

Proposed System

The proposed Indian Crop Yield Analysis and Prediction system provides an integrated platform for analyzing agricultural data and predicting crop yield. Users can upload historical crop datasets containing information about crops, regions, cultivated area, production, weather conditions, and other agricultural variables. The system validates the dataset and performs preprocessing to handle missing values, duplicate records, inconsistent units, and invalid entries. Feature engineering creates useful attributes required for analytical and Machine Learning operations. The analytics module calculates production, yield, year-wise growth, crop-wise performance, and regional statistics. Users can compare crop yield across different states, districts, years, and crop categories. Environmental analysis can examine relationships between rainfall, temperature, soil properties, irrigation, and crop yield. Interactive visualizations present trends through charts, graphs, maps, tables, and KPI cards. The Machine Learning module trains a regression model using historical agricultural data and selected features. The prediction module accepts new agricultural inputs and generates an estimated crop-yield value. Model performance is evaluated using appropriate regression metrics before deployment. The proposed system therefore combines agricultural analytics and Machine Learning-based yield prediction in a centralized platform.

Advantages of Proposed System

- Provides centralized crop-yield analysis.
- Automates agricultural data preprocessing.
- Supports crop-wise and region-wise analysis.
- Identifies historical yield trends.
- Supports year-wise production comparison.
- Analyzes rainfall and temperature relationships.
- Can incorporate soil and irrigation information.
- Provides interactive charts and dashboards.
- Uses Machine Learning for yield prediction.
- Reduces manual data-analysis effort.
- Supports data-driven agricultural planning.
- Provides a foundation for future precision-agriculture features.

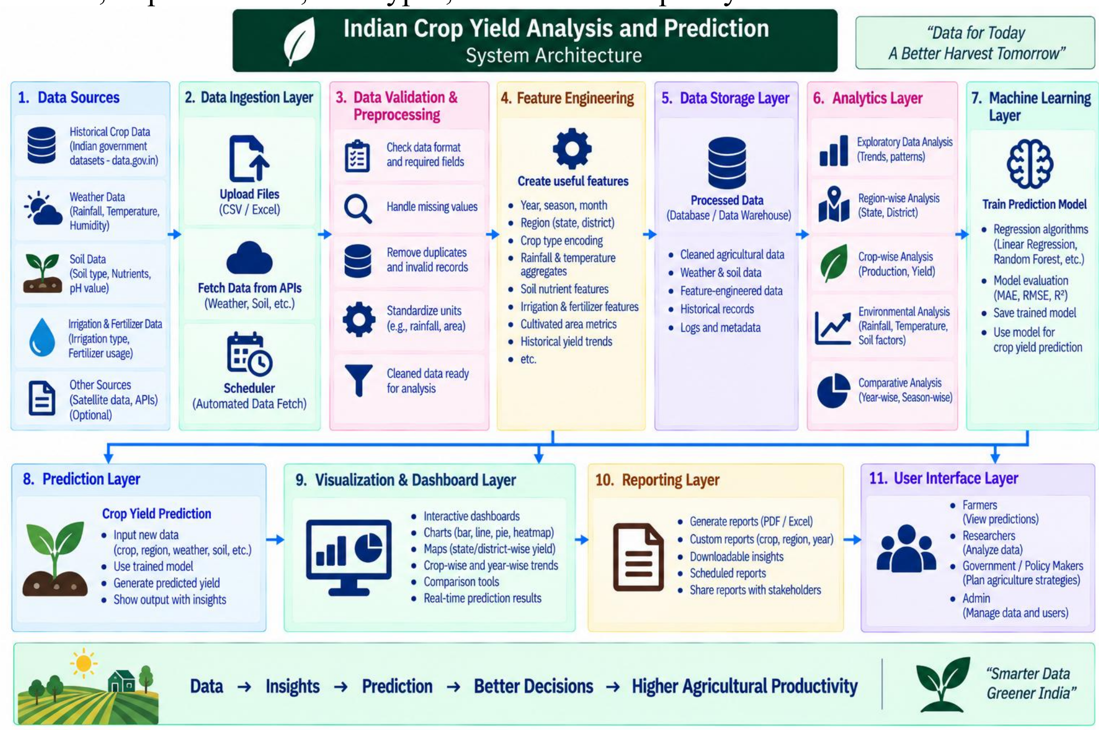
IV. METHODOLOGY

The methodology begins with collecting historical Indian agricultural data from authorized datasets or agricultural databases. The collected data is validated to check its structure, required attributes, data types, and completeness. Data preprocessing is performed to remove duplicate records and handle missing, invalid, and inconsistent values. Numerical features such as cultivated area, production, rainfall, and temperature are standardized when required. Categorical variables such as state, district, and crop type are transformed into suitable formats for analysis and Machine Learning. Feature engineering is performed to create useful attributes related to time, region, crop type, and environmental conditions. Exploratory Data Analysis is conducted to identify crop production and yield patterns. Comparative analysis is performed across states, districts, crop types, and different years. Visualization techniques are used to represent trends through bar charts, line graphs, maps, scatter plots, and dashboards. Relevant agricultural features are selected for predictive modeling based on data quality and analytical relevance. A regression-based Machine Learning model is trained to predict crop yield from historical agricultural features. The trained model is evaluated using suitable metrics such as MAE, MSE, RMSE, and R^2. Finally, the system displays analytical insights and predicted crop-yield values through an interactive user interface.

SYSTEM ARCHITECTURE

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

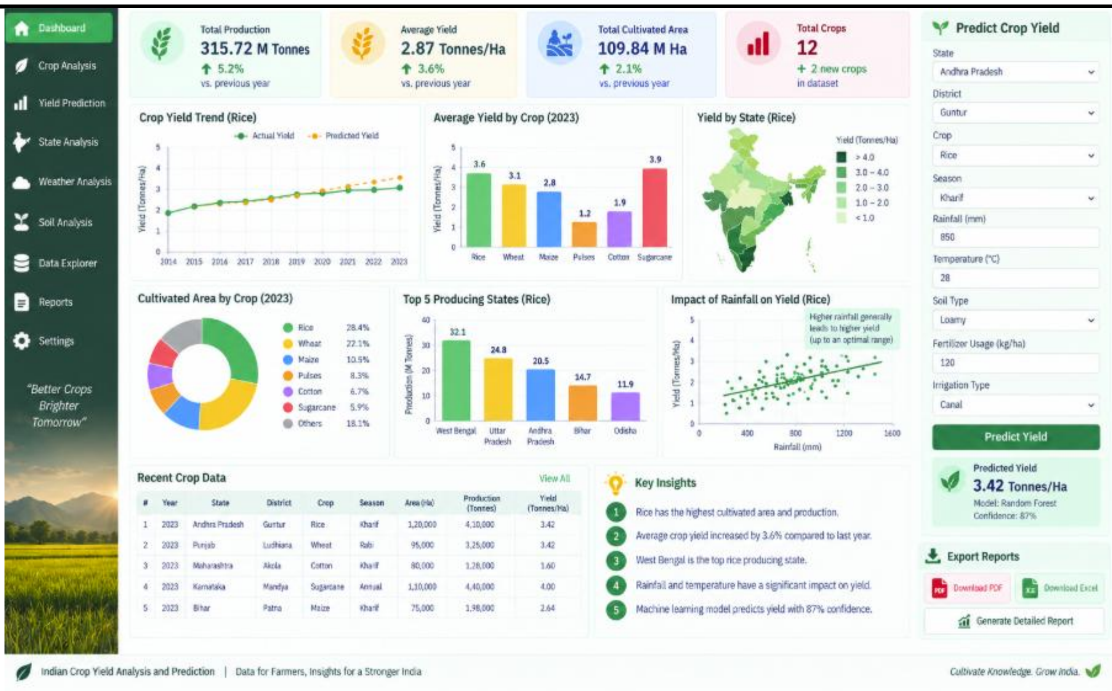
The system architecture consists of several interconnected layers that process agricultural data and generate crop-yield insights and predictions. The Data Source Layer provides historical crop records, weather information, soil data, irrigation information, and other authorized agricultural datasets. The Data Input Layer accepts data through CSV, Excel, databases, or APIs. The Data Validation Layer checks file formats, required fields, data types, and basic data quality.



The Data Preprocessing Layer handles missing values, duplicate records, invalid entries, and inconsistent units. The Feature Engineering Layer creates useful attributes such as crop type, region, year, rainfall range, temperature range, and cultivated-area features. The Data Storage Layer securely stores cleaned agricultural records and historical information. The Analytics Processing Layer calculates crop production, yield, growth trends, and regional statistics. The Regional and Crop Analysis Layer compares crop performance across states, districts, crop types, and years. The Environmental Analysis Layer examines relationships between rainfall, temperature, soil properties, irrigation, and crop yield. The Machine Learning Layer trains and evaluates a suitable regression model using historical agricultural data. The Prediction Layer accepts new agricultural inputs and generates an estimated crop-yield value. The Visualization and Reporting Layer presents analytical results, prediction outputs, charts, maps, tables, and KPI cards through an interactive dashboard. The User Interface Layer allows authorized users such as analysts and agricultural researchers to explore data and prediction results.

V. RESULT AND OUTPUT

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



VI. CONCLUSION

The Indian Crop Yield Analysis and Prediction system provides a data-driven solution for analyzing agricultural data and predicting crop yield. The system combines data preprocessing, exploratory data analysis, visualization, and Machine Learning techniques to convert historical agricultural records into useful insights.

The system analyzes important factors such as crop type, cultivated area, production, rainfall, temperature, soil conditions, irrigation, fertilizer usage, state, district, and year. This helps identify crop-wise, region-wise, and time-based patterns in agricultural productivity.

The dashboard presents important information through KPI cards, charts, graphs, maps, tables, and prediction results. Users can easily examine crop production trends, compare yields across regions, analyze environmental factors, and understand variations in agricultural performance.

The Machine Learning component provides an estimated crop yield based on the selected agricultural and environmental inputs. This predictive capability can support agricultural researchers, analysts, farmers, and administrators in understanding potential yield outcomes and planning agricultural activities.

In the future, the system can be enhanced with real-time weather data, satellite imagery, IoT soil sensors, crop disease detection, irrigation recommendations, fertilizer optimization, advanced Machine Learning models, and real-time yield forecasting. These enhancements can make the system more intelligent and useful for data-driven and sustainable agricultural planning.

REFERENCES

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.