Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/29
DOI: https://doi.org/10.65713/jdebmv2i129

CROP YIELD PREDICTION USING FEATURE SELECTION TECHNIQUES AND ENSEMBLE MACHINE LEARNING CLASSIFIERS

[#1]**CHERUVU SONIA ANUSHA,** ***M.Tech(SE) Student,***
[#2]**Dr. M. ANJAN KUMAR,*****Professor, Dept of CSE,***
[#3]**Mrs. T.SUBBALAKSHMAMMA,** ***Assistant Professor, Dept of CSE,***
VISWAM ENGINEERING COLLEGE(AUTONOMOUS), MADANAPALLE, AP.

ABSTRACT: In order to analyze the characteristics of the agricultural environment, this study focuses on crop forecasting using a variety of feature selection techniques and classifiers. Accurate crop projections are crucial for increasing agricultural productivity, optimizing resource utilization, and developing sustainable farming practices. In order to determine how environmental factors affect crop adaptability, the study looks at soil nutrients, temperature, humidity, rainfall, and pH levels. The most important elements influencing crop growth are identified using a variety of feature selection techniques, which improves forecast accuracy and reduces data dimensionality. To create predictive models and evaluate their efficacy, a range of machine learning classifiers are used, including decision trees, support vector machines, k-nearest neighbors, and random forest techniques. The experimental findings show that combining effective feature selection techniques with robust classification algorithms significantly improves crop prediction accuracy. This helps farmers and agricultural planners make well-informed decisions on crop cultivation.

Keywords:*Crop Prediction, Agricultural Environment, Feature Selection Techniques, Machine Learning Classifiers, Soil Nutrients, Temperature, Humidity,*

1. INTRODUCTION

Agriculture is crucial to maintaining the world economy and ensuring food security, particularly in developing countries where farming is the main source of income. As the population grows, so does the need for food, which calls for the use of cutting-edge technologies to boost agricultural output. Crop forecasting is currently a key component of intelligent agriculture. It helps farmers and agricultural planners determine which crops would thrive in particular climates. Accurate crop forecasting not only increases food output but also maximizes the use of natural resources including soil nutrients, water, and fertilizers.

A number of variables, such as soil composition, temperature, precipitation, humidity, and pH levels, can affect the agricultural environment. All of these factors have a significant impact on crop growth and yield. In the past, farmers chose crops based on their personal experience and historical knowledge, but these methods were not always successful because of shifting weather patterns and other environmental conditions. Machine learning algorithms can now efficiently analyze large agricultural datasets and evaluate the effects of numerous environmental conditions on crop adaptation thanks to developments in data-driven technology.

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/29
DOI: https://doi.org/10.65713/jdebmv2i129

However, agricultural databases frequently include a wide range of characteristics, some of which can be unnecessary or unimportant. Predictive algorithms may become less accurate and useful as a result of such data. Because they pinpoint the crucial elements influencing crop development, feature selection techniques are crucial for resolving this problem. By eliminating unnecessary features, these techniques improve model performance, lower data dimensionality, and speed up computations. Numerous techniques, such as filter-based, wrapper-based, and embedding methodologies, can be used to identify important environmental factors for crop prediction.

To create precise crop prediction models, a variety of machine learning classifiers are used in addition to feature selection. In agricultural data processing, algorithms such as Decision Trees, Random Forest, Support Vector Machines, Naïve Bayes, and K-Nearest Neighbors have proven to be effective. Before determining the best crops for the specified parameters, these classifiers look for trends in environmental datasets. Combining feature selection techniques with machine learning classifiers improves predicted accuracy by using just the best variables for model training.

2. LITERATURE SURVEY

Anderson et al. (2025): In order to determine the most suitable crops for a variety of agricultural regions, we introduce a crop prediction framework that is based on machine learning and takes into account the environmental parameters of those regions. Environmental information from agricultural datasets, including soil pH, temperature, precipitation, humidity, and soil nutrients, is incorporated into the study. Several feature selection techniques, including Information Gain and Chi-square, identify important factors influencing agricultural output. To determine which crops are best suited for cultivation, we employ classification techniques such as Random Forest and Support Vector Machine. According to experimental findings, choosing relevant environmental features improves prediction accuracy and makes it easier to make well-informed agricultural decisions.

Gupta & Sharma (2024): A data-driven agricultural forecasting system that uses environmental elements to choose suitable crop varieties is presented in this research. The framework works with datasets that include soil fertility, climate, and moisture levels. Two techniques used to increase model effectiveness and simplify data are Principal Component Analysis and Correlation-based Feature Selection. Agricultural yields are predicted using a variety of classifiers, including Decision Tree and Naïve Bayes. The findings show that when feature selection is combined with machine learning classifiers, crop suggestions are noticeably more accurate.

Lee & Park (2023): The use of machine learning to generate intelligent crop forecasts based on agricultural environmental features is examined in this work. To determine which crops will thrive, the method assesses soil composition, temperature trends, rainfall patterns, and humidity levels. To find the most important environmental features, we employ feature selection techniques like Recursive Feature Elimination. Predictive models are built using classification methods such as k-Nearest Neighbors and Logistic Regression. The findings show

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/29
DOI: https://doi.org/10.65713/jdebmv2i129

that improved feature selection improves classifier performance and makes agricultural planning easier.

Rodriguez et al. (2022): Environmental data is used by a predictive agricultural analytics tool to determine the best crops for different farming environments. The study makes use of databases that contain data on soil characteristics, weather patterns, and irrigation settings. To identify the key factors influencing crop development, we employ feature selection techniques such wrapper-based approaches and mutual information. Crop forecasting uses two different kinds of classification algorithms: Random Forest and Gradient Boosting. Studies show that machine learning techniques can effectively help farmers choose crops that are appropriate for their surroundings.

Kumar & Patel (2021): This study investigates the utilization of machine learning classifiers for crop prediction using agricultural environmental datasets. We analyze factors including soil concentrations of nitrogen, phosphate, and potassium as well as precipitation, temperature, and humidity levels to determine how well crops develop. Two feature selection techniques that help identify the crucial elements influencing agricultural productivity are ReliefF and Genetic Algorithms. Predictive models are built using classification techniques such as Support Vector Machine and Decision Tree. The findings show that combining feature selection and classification methods significantly improves crop prediction accuracy and promotes sustainable farming methods.

3.RELATED WORK

Soil Condition Based Crop Prediction

The topography influences food production and planning. Experts consider moisture, pH, and nutrient levels to determine which crops flourish in particular soil types. Farmland is classified based on soil quality through machine learning and image analysis. These approaches assess the color, texture, and mineral composition of the soil to find the most productive crops. Computerized soil analysis can assist farmers in enhancing agricultural output. These instruments aid farmers in crop selection, soil management, and nutrient enhancement.

Environmental Condition Based Crop Prediction

Precipitation, wind, solar radiation, humidity, and temperature all influence agricultural development and productivity. Contemporary crop forecasting systems utilize agricultural sensors, satellites, and meteorological data to assess weather conditions and food production. Satellite imagery and other remote sensing technologies are employed in industrial agriculture to monitor plant development. Forecasting models can aid farmers in planning for droughts and floods by evaluating historical patterns and projected weather changes.

Fig1. Multi-Source Remote Sensing Framework for Smart Agriculture

Remote Sensing Based Crop Monitoring

Farmers increasingly utilize remote sensing to assess crop health and predict harvest outcomes. Aerial photography

JDEBM Journal of Digital Engineering and Business Managemet

Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/29
DOI: https://doi.org/10.65713/jdebmv2i129

reveals crop health, plant development, and soil moisture levels throughout extensive agricultural areas. These images enable the identification of diseases, crop stress, and environmental alterations through advanced data processing techniques. Remote sensing enables academics and farmers to assess agricultural productivity without the necessity of personally visiting the location. This system enhances farm planning by delivering rapid and accurate crop data, thereby enabling the detection of production inefficiencies.

Machine Learning Techniques for Crop Prediction

Machine learning predicts agricultural crop production utilizing extensive datasets. These methodologies predict agricultural growth by analyzing soil conditions, meteorological data, and past crop yields. Machine learning algorithms can forecast agricultural output via autonomous data analysis. Agricultural professionals can employ these strategies to ascertain optimal growing conditions for plants. Agricultural managers can optimize fertilizer and water utilization with machine learning.

Classification Techniques in Crop Prediction

Classification algorithms are employed by numerous crop forecasting tools to categorize agricultural data. These strategies ascertain the optimal diet for a location by evaluating soil nutrients, humidity, precipitation, and temperature. Data classification algorithms encompass support vector machines, random forests, decision trees, and k-nearest neighbors. These algorithms ascertain the ideal crop yields for each plot through the analysis of agricultural trends. Agriculturists can choose regional crops and assess crop performance via classification algorithms.

Smart Agriculture and IoT Based Monitoring

Smart agricultural machinery can assess crops and their environment through the Internet of Things. Agricultural sensors assess temperature, humidity, soil moisture, and nutrient concentrations. The data is then transmitted to computers for machine learning purposes. Data-driven agricultural solutions can enhance food yield, diminish fertilizer consumption, and automate irrigation processes. The outcomes are increased agricultural productivity and diminished labor expenses.

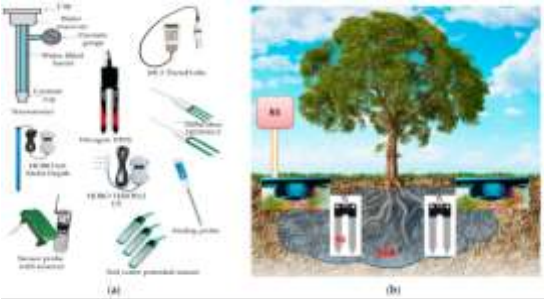

Fig2. Soil Moisture and Soil Water Potential Sensor System for Plant Monitoring

4.RESULTS

Fig4.1 User Login

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/29
DOI: https://doi.org/10.65713/jdebmv2i129

Fig4.2 Register Your Details here

Fig4.3 View all remote users


Fig4.5 Enter all crop Datasets Details


Fig4.6 Crop Details trained and tested results


Fig4.7 Bar graph


Fig4.7 Line chart


Fig4.7 Pie chart

5. CONCLUSION

Machine learning classifiers and feature selection algorithms may accurately estimate crop yields, enhancing agricultural production and decision-making. Predictive models consider soil nutrients, pH, precipitation, humidity, and climatic conditions while selecting crops. Feature selection strategies facilitate the identification of valuable features, the reduction of data dimensionality, and the improvement of classification systems. These methodologies forecast agricultural yield with machine learning algorithms. These advancements improve agricultural

JDEBM Journal of Digital Engineering and Business Managemet
Volume-2, Issue-1, January-March 2026
ISSN: 3107-7897 | P.ID: JDEBM/29
DOI: https://doi.org/10.65713/jdebmv2i129

decision-making, aiding farmers in resource conservation and prolonging farm viability.

REFERENCES

[1] R. Jahan, ‘‘Applying naive Bayes classification technique for classificationof improved agricultural land soils,’’ Int. J. Res. Appl. Sci. Eng. Technol.,vol. 6, no. 5, pp. 189–193, May 2018.

[2] B. B. Sawicka and B. Krochmal-Marczak, ‘‘Biotic components influencing the yield and quality of potato tubers,’’ Herbalism, vol. 1, no. 3,pp. 125–136, 2017.

[3] B. Sawicka, A. H. Noaema, and A. Gáowacka, ‘‘The predicting the sizeof the potato acreage as a raw material for bioethanol production,’’ inAlternative Energy Sources, B. Zdunek, M. Olszáwka, Eds. Lublin, Poland:WydawnictwoNaukowe TYGIEL, 2016, pp. 158–172.

[4] B. Sawicka, A. H. Noaema, T. S. Hameed, and B. Krochmal-Marczak,‘‘Biotic and abiotic factors influencing on the environment and growthof plants,’’ (in Polish), in Proc. BioróżnorodnośćŚrodowiskaZnaczenie, Problemy, Wyzwania. MateriałyKonferencyjne, Puławy, May 2017.[Online]. Available: https://bookcrossing.pl/ksiazka/321192

[5] R. H. Myers, D. C. Montgomery, G. G. Vining, C. M. Borror, andS. M. Kowalski, ‘‘Response surface methodology: A retrospective andliterature survey,’’ J. Qual. Technol., vol. 36, no. 1, pp. 53–77, Jan. 2004.

[6] D. K. Muriithi, ‘‘Application of response surface methodology for optimization of potato tuber yield,’’ Amer. J. Theor. Appl. Statist., vol. 4, no. 4,pp. 300–304, 2015, doi: 10.11648/j.ajtas.20150404.20.

[7] M. Marenych, O. Verevska, A. Kalinichenko, and M. Dacko, ‘‘Assessmentof the impact of weather conditions on the yield of winter wheat inUkraine in terms of regional,’’ Assoc. Agricult. Agribusiness Econ. Ann.Sci., vol. 16, no. 2, pp. 183–188, 2014.

[8] J. R. Olędzki, ‘‘The report on the state of remotesensing in Poland in2011–2014,’’ (in Polish), Remote Sens. Environ., vol. 53, no. 2,pp. 113–174, 2015.

[9] K. Grabowska, A. Dymerska, K. Poáarska, and J. Grabowski, ‘‘Predictingof blue lupine yields based on the selected climate change scenarios,’’ ActaAgroph., vol. 23, no. 3, pp. 363–380, 2016.

[10] D. Li, Y. Miao, S. K. Gupta, C. J. Rosen, F. Yuan, C. Wang, L. Wang, andY. Huang, ‘‘Improving potato yield prediction by combining cultivar information and UAV remote sensing data using machine learning,’’ RemoteSens., vol. 13, no. 16, p. 3322, Aug. 2021, doi: 10.3390/rs13163322.

[11] N. Chanamarn, K. Tamee, and P. Sittidech, ‘‘Stacking technique for academic achievement prediction,’’ in Proc. Int. Workshop Smart Info-MediaSyst., 2016, pp. 14–17.

[12] W. Paja, K. Pancerz, and P. Grochowalski, ‘‘Generational feature elimination and some other ranking feature selection methods,’’ in Advancesin Feature Selection for Data and Pattern Recognition, vol. 138. Cham,Switzerland: Springer, 2018, pp. 97–112.