Crop Production in Agriculture

SAI KUMAR MURARISHETTI
SAIKUMARMURARISHET@LEWISU.EDU
LEWIS UNIVERSITY

Abstract: The current studies analyze 28 years of data to examine various factors such as environment factors, different types of crops and the results of agriculture in India. The government resources were used to construct the datasets, containing important information including temperature, precipitation, pH levels, area, state-by-state data, and agricultural production details. The project gives predicted information about agricultural productivity through the implementation of supervised machine learning algorithms, providing crop management and well-informed decision-making. Overall highlight the value of using techniques based on data for better resource allocation and assists India's agriculture industry. When it comes to using data for sustainable agriculture practices, this investigation is an important asset for investigators, decision-makers, and researchers.

Keywords: Logistic Regression model, Decision Tree.

I. Introduction (*Heading 1*)

Within the field of artificial intelligence (AI), machine learning is the branch that focuses on creating models and algorithms that enable computers to learn and make predictions or judgments without the need for programming. It's a fast-moving field with many of applications across various sectors. There are several categories in machine learning some are Supervised machine learning, Unsupervised machine learning and Deep Learning and many more.

In Supervised Machine Learning the algorithm learns to make predictions or classifications based on past data or labeled training data. In supervised learning, the algorithm is provided with a dataset that contains both input features and corresponding output labels. The algorithm's objective is to understand the relationship among the inputs and outputs so that it may correctly classify or predict new, unidentified data.

By applying Supervised Machine Learning to the model data will classify crops into different categories such as high-yield, low-yield, disease-resistant, or crop type based on features such as temperature, rainfall, pH levels, and state. It will help in crop management and selection and datasets are ideal for regression problems, it can predict crop production (in tons) or yield per hectare based on various input features. For example, it can predict rice production based on temperature, rainfall, and pH levels.

The dataset used in this research was obtained from Kaggle.com.

II. Description

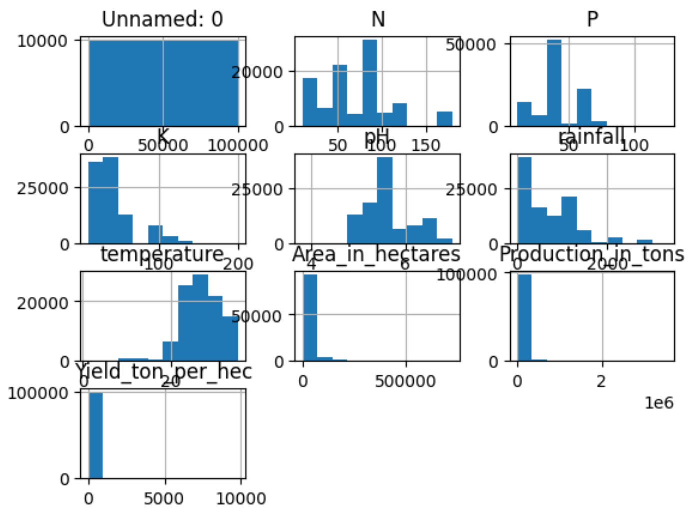
Supervised machine learning uses parameters like temperature, rainfall, pH levels, and condition to classify crops into different categories, such as high-yield, low-yield, or based on disease resistance and crop varieties. This process provides important insights for crop management and selection. The datasets are suitable for regression analysis as they allow yield per hectare or crop output in tons to be predicted using historical data and a few specific characteristics, including temperature, rainfall, and pH levels. Machine learning models can be trained to forecast crop yields by utilizing variables like temperature, rainfall, pH levels, and other relevant features. To examine its effect in changes of weather patterns on crop yield while considering the 28 years of the datasets. These databases are necessary for the agricultural sector's risk assessment, planning of resources, data-driven decision-making, and crop yield predictions.

III. Methodology

Crop Yield Forecasting: The capacity of well-trained model to forecast crop output in the future are necessary for farmers, public officials and the agricultural sector, since they allow for better resource allocation and decision-making.

Optimization and Decision Support: Supervised machine learning models are essential for improving resource allocation, crop selection and pest management methods. By using past data and expected results, these models support decision-making and increase agricultural efficiency and sustainability.

Updates and Monitoring: A system for continuing updates and model monitoring is required for ensuring the relevance and accuracy of predictive models since the frequently changing character of agricultural data.



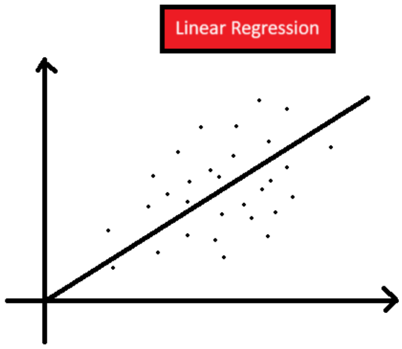
The data. hist() method mentioned typically generates histograms of the features or attributes in Crop_Production dataset.

IV. Methods of Supervised Machine Learning

A. Linear Regression:

Through the fitting of a linear equation, linear regression describes the connection between a dependent variable and one or more independent variables. Predicting continuous numerical values is one of its frequent applications.

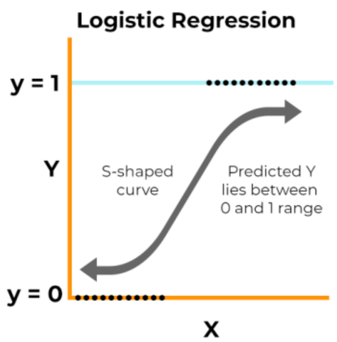
Real-time Example: Determining the cost of a home by taking into account variables like location, number of bedrooms, and square footage. For instance, given these features, the model can estimate the selling price of a house.



B. Logistic Regression:

For binary classification issues, logistic regression is utilized. It determines the likelihood that an input, usually 0 or 1, belongs to a certain class.

Real-time Example: Predicting whether an email is spam or not based on its content and characteristics. The model assigns a probability to each email, and if the probability is above a certain threshold, it classifies the email as spam.



The accuracy of the logistic regression model from the Crop_Production dataset.

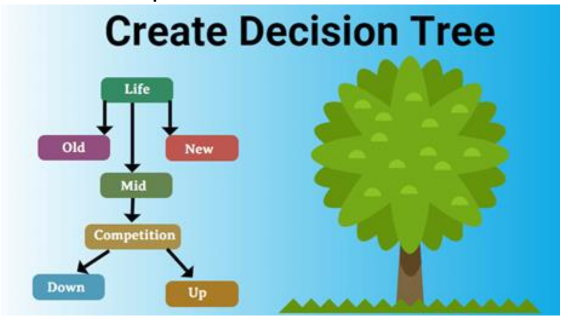
Logistic Regression Model:
Classification Report:

	precision	recall	f1-score	support
0	0.85	0.87	0.86	7640
1	0.92	0.96	0.94	5577
2	0.85	0.72	0.78	1400
3	0.87	0.82	0.84	5353
accuracy			0.87	19970
macro avg	0.87	0.84	0.85	19970
weighted avg	0.87	0.87	0.87	19970

Accuracy Score: 0.8726089133700551

C. Decision Trees:

Decision trees are versatile for issues involving both regression and classification. They create a tree-like structure of decisions and consequences based on input features.



Real-time Example: Deciding whether to go for a picnic based on weather conditions. The decision tree considers factors like temperature, humidity, and wind speed to make the decision, providing a clear path to the choice.

The accuracy of the Decision Tree model from the Crop_Production dataset.

[3 1 0 ... 3 0 3]
Classification Report:

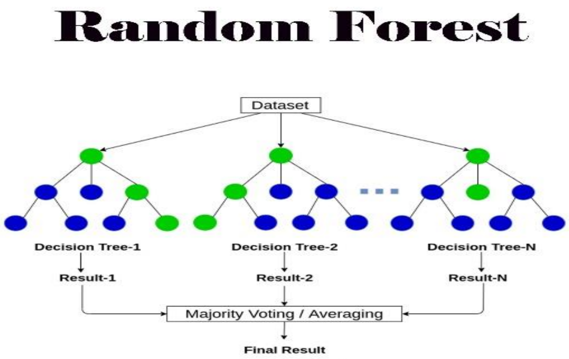
	precision	recall	f1-score	support
0	1.00	1.00	1.00	7640
1	1.00	1.00	1.00	5577
2	1.00	1.00	1.00	1400
3	1.00	1.00	1.00	5353
accuracy			1.00	19970
macro avg	1.00	1.00	1.00	19970
weighted avg	1.00	1.00	1.00	19970

Accuracy Score: 0.38257386079118677

D. Random Forest:

Several decision trees are combined in Random Forest, an ensemble learning technique, to improve prediction accuracy and decrease overfitting.

Real-time Example: Predicting customer churn in a telecom company. The Random Forest model takes into account various customer attributes and behaviors to determine the likelihood of a customer canceling their subscription.



The accuracy of the Random Forest model from the Crop_Production dataset.

Random Forest Model:
Classification Report:

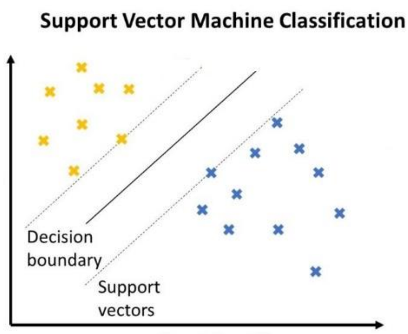
	precision	recall	f1-score	support
0	1.00	1.00	1.00	7640
1	1.00	1.00	1.00	5577
2	1.00	1.00	1.00	1400
3	1.00	1.00	1.00	5353
accuracy			1.00	19970
macro avg	1.00	1.00	1.00	19970
weighted avg	1.00	1.00	1.00	19970

Accuracy Score: 0.999899849774662

E. Support Vector Machines (SVM):

For binary classification issues, SVM is employed. It locates the hyperplane that maximizes the margin between two classes and best divides them apart.

Real-time Example: Classifying images as either cats or dogs based on their features. The SVM algorithm determines the optimal decision boundary to distinguish between the two classes.



V. CONCLUSION

The comprehensive analysis of 28 years of agricultural data in India, combined with application of supervised machine learning algorithms, sheds light on the multifaceted relationship between environmental factors, crop types, and agricultural outcomes. These insights offer a valuable resource for researchers, decision-makers, and the agriculture industry, promoting data-driven practices and informed decision-making.

Supervised machine learning, a powerful subset machine learning, proves instrumental in harnessing the potential of this rich dataset. The utilization of parameters such as temperature, rainfall, pH levels, and crop conditions enables the classification of crops into various categories. Whether it's distinguishing between high-yield and low-yield crops or identifying disease-resistant varieties, these models empower precision in crop management and selection.

Furthermore, supervised machine learning extends its capabilities to regression analysis, providing predictive prowess. By examining the interplay of factors like temperature, rainfall, and pH levels, these models forecast crop production in tons or yield per hectare. For instance, they can accurately predict rice production based on weather variables, offering invaluable information for resource allocation and agricultural planning.

The dataset, sourced from Kaggle.com, underpins the potential of these techniques. As the datasets span 28 years, they afford an opportunity to study the dynamic influence of changing weather patterns on crop yield. This temporal analysis is essential for risk assessment, resource planning, and the formulation of data-driven strategies, including crop yield predictions.

Looking ahead, the impact of supervised machine learning in agriculture is profound. Crop yield forecasting, driven by well-trained models, holds the key to better resource allocation for farmers, policymakers, and the agricultural sector. Moreover, it plays an integral role in optimization and decision support, enhancing the efficiency and sustainability of agricultural practices.

In the dynamic realm of agriculture, where data continually evolves, regular updates and model monitoring are pivotal. These processes ensure that predictive models remain accurate and relevant, aligning with the ever-changing landscape of agricultural data.

In summation, this research project underscores the significance of integrating supervised machine learning into agriculture, presenting an innovative approach to enhance crop management, risk assessment, and decision-making. The invaluable insights garnered from this endeavor offer a blueprint for a sustainable and data-driven future for India's agriculture sector.

The journey of data-driven agriculture is underway, and the path to optimized resource allocation, improved yields, and informed choices is illuminated by the synergy between agriculture and supervised machine learning.

REFERENCES

[1] Pan, W.-T. A new fruit fly optimization algorithm: Taking the financial distress model as an example. Knowl.-Based Syst. 2012, 26, 69–74

[2] Madhumathi, R.; Arumuganathan, T.; Shruthi, R. Soil NPK and Moisture analysis using Wireless Sensor Networks. In Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 1–3 July 2020.

[3] Crop Yield Prediction Using Machine Learning Algorithms, Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal,2019 Fifth International Conference on Image Information Processing (ICIIP), pp 125-130

[4] Renuka & Sujata Terdal, "Evaluation of Machine Learning Algorithms for Crop Prediction" Published in International Journal of Engineering and Advanced Technology (IJEAT) Volume-8 | August 2019.

[5] Andrew Crane Droesch, “Machine learning methods for crop yield

prediction and climate change impact assessment in agriculture”,
Published by IOP Publishing Ltd
Volume: 05 | OCT -2018.
[6] Nischitha K, Dhanush Vishwakarma, Mahendra N, Ashwini, Manjuraju M R, “Crop Prediction using Machine Learning Approaches”,
International Journal of Engineering Research and Technology, vol.9
Issue 08, August-2020 ISSN: 2278-0181.
[7] Reddy, D. Anantha, BhagyashriDadore, and Aarti Watekar. “Crop recommendation system to maximize crop yield in ramtek region using machine learning.” International Journal of Scientific Research in Science and Technology 6.1 (2019): 485-489.
[8] https://www.kaggle.com/datasets/asishpandey/crop-production-in-india/data//