Grenze International Journal of Engineering and Technology, June Issue

Machine Learning Model for Agriculture Crop: A Review Study

Saurabh Shrivastava[1], Ramnaresh Sharma[2], Dr. Pritaj Yadav[3] and Dr. Jitendra Singh Kushwah[4]
[1-3] Rabindranath Tagore University Bhopal-MP-India
Email: shrivastava901@gmail.com, ramcse1983@gmail.com, pritaj.yadav@aisectuniversity.ac.in
[4]Institute of Technology and Management Gwalior-MP- India
Email: jitendra.singhkushwah@itmgoi.in

***Abstract*—In contrast to the former system, which only took into account one state, we used a big dataset that covered all of India's states in our new method. These recommendations could be taken out and applied to teach the farmers. By creating a visual representation, the farmer can better comprehend the crops to grow. With the use of machine learning techniques, we can make predictions and create a clear model from the data. It is possible to resolve agricultural problems such crop rotation, crop prediction, water and fertiliser needs, and crop protection. In contrast to the former system, which only took into account one state, we used a big dataset that covered all of India's states in our new method. These recommendations could be taken out and applied to teach the farmers. By creating a visual representation, the farmer can better comprehend the crops to grow. With the use of machine learning techniques, we can make predictions and create a clear model from the data. It is possible to resolve agricultural problems such crop rotation, crop prediction, water and fertiliser needs, and crop protection. Crops are suggested for use in such a strategy depending on their quantity and meteorological considerations. Data analytics opens the door for the development of valuable agricultural database extraction. After analysing the crop data set, crops are recommended based on their productivity and the season.**

***Index Terms*— Machine learning, Agriculture techniques, crop predictions.**

I. INTRODUCTION

Changing temperature, seasons, and soil conditions all contribute to agricultural unpredictability, which lowers production. A higher population and area should increase productivity, yet it is impossible to do so. Farmers once relied on word-of-mouth, but they are unable to do so due to climate issues. Agricultural aspects and features are utilised to give data that may be used to learn more about Agri-facts. Agriculture Sciences provides important IT industry highlights to help farmers with reliable agriculture information. The ability to apply new technical methods to the agricultural industry is desirable given the current circumstances. Using data, machine learning techniques build a precise model that may be applied to prediction. Problems with crop forecasting, rotation, water and fertiliser needs, and crop protection might all be solved. The brain of machine learning is where all learning occurs. The machine learns in a similar manner to how people do. People learn by their experiences. Making forecasts is made simpler the more data we have. By analogy, a situation where the outcome is unknown has fewer chances of success than one where it is. The same methods are used to teach machines. To get a precise forecast, the system looks at an example. The machine is capable of making

Grenze ID: 01.GIJET.9.2.387

predictions [1] when we present it with similar circumstances. But, just like a person, the computer struggles to predict what will happen next if it is given an example that has never happened before.
The two basic objectives of machine learning are inference and learning. The machine learns mostly through identifying patterns.
Programs for machine learning have a distinct life cycle, which can be summarised as follows:

1) Formulate a query

2) Gather information

3) Visualize data

4) Train the algorithm

5) Put the algorithm to the test

6) Gather feedback

7) Fine-tune the algorithm

8) Repeat steps 4–7 until the results are satisfactory.

9) Make a forecast using the model.

The algorithm applies what it has learned to new data sets once it has mastered drawing the correct conclusions. Technologies like blockchain, IoT, machine learning, deep learning, cloud computing, and edge computing can be utilised to gather and process data. Applications for IoT, computer vision, and machine learning will help farmers and other associated industries improve output, quality, and profitability[2].
The three main types of agricultural activities are pre-harvest, harvest, and post-harvest. Advances in machine learning have contributed to increased agricultural productivity. Machine learning, a relatively new technology, helps farmers cut down on farming losses by providing thorough crop recommendations and insights. The most beneficial and cutting-edge approach that is advised for usage in agriculture-based applications is machine learning. By enrolling into a personalised dashboard on a computer or tablet and using machine learning technology, a farmer can receive an accurate estimate of the number of acres that are suitable for harvesting on a given day. It is possible to estimate and project the weight and maturity of harvestable crops. Also, utilising a variety of technologies, such as image analysis, crops can be assessed both before and after harvest for the presence of desirable features, level of damage (if applicable), nutritional make-up, and other factors that may affect the final viable yield and product pricing. Many industries employ machine learning (ML) techniques, from forecasting consumer phone usage to analysing supermarket patron behaviour. Agriculture has employed machine learning for some time. One of the most difficult problems in precision agriculture is predicting crop production, and numerous models have been suggested and tested up to this point. Since the production of crops is influenced by a variety of elements, including climate, weather, soil, the usage of fertilisers, and seed type, this task requires the use of multiple datasets[3]. This suggests that measuring agricultural output requires a series of difficult stages rather than a straightforward method. Although improved yield prediction ability is being sought after, crop yield prediction programmes are currently able to reasonably anticipate actual yields. The issue of how climate change affects agricultural productivity has been resolved by researchers in a number of ways.
In order to identify the topic of further research, we endeavour to critically analyse the work of colleagues and offer a brief assessment of it in this part. As part of our research, we shall carry on with our investigation. Four broad categories can be used to group its applications:
i) Crop management
ii) Livestock management
iii) Water management
iv) Soil management

II. Related Work

A technique was developed by L.Ismail to forecast a nation's crisis preparation. Machine learning is being used to study climate change. The study's region of emphasis is Southeast Asia. Data gathering, data training, data testing, and index computing are the steps taken in the predictive index calculation process. Among other things, an index can be used for prediction, index validation, and index visualization[4]. The research is being done as a preventative precaution. The regions will be informed, and deep learning will be used to check the vulnerable index.

III. Proposed Work

For decision-makers at all levels, including those in national and regional (like the EU) decision-making, crop yield forecasting is a significant difficulty. With an accurate crop production forecast model, farmers can help them choose what to produce and when. For decision-makers at all levels, including those at the national and regional (such as EU) levels, predicting crop yields can be challenging. A crop production prediction model is a tool that farmers can use to decide what and when to sow. Forecasting crop production can be done using a variety of techniques. This review study examines the research on the application of machine learning to agricultural yield prediction. One of the exclusion criteria in our analysis of the recovered publications is whether the publication is a survey or a conventional review piece. The items that were omitted are actually a component of a bigger undertaking that is covered in this section.

IV. Machine Learning Implementation

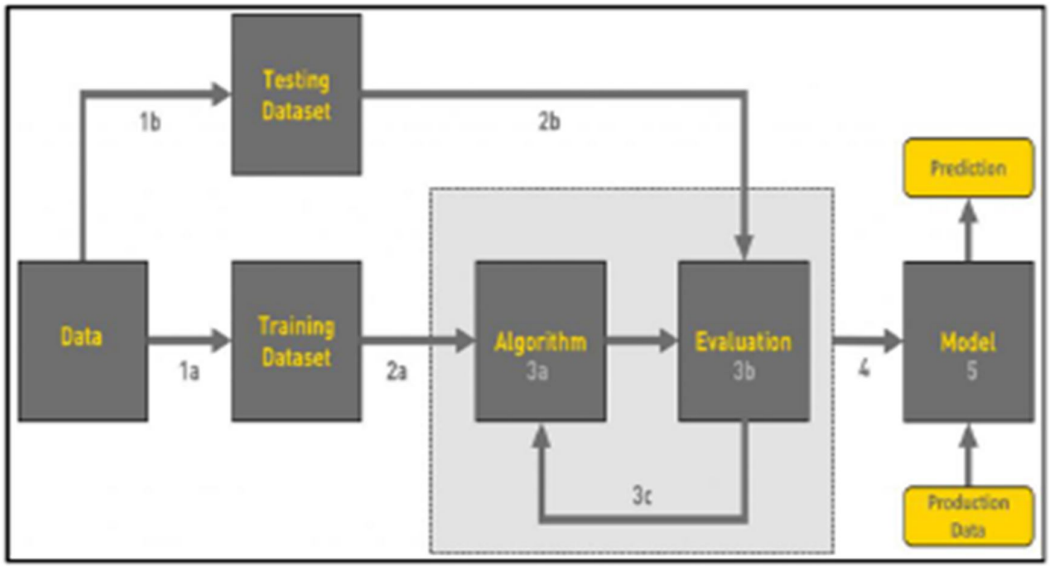

Fig. 1. Machine Learning Block Diagram

Libraries for Python are needed for implementation
i) Numpy
ii) Pandas
iii) Sci-kit Learn
iv) Matplotlib

Pre-processing of Data
The type of project we want to construct will influence the data collection method we use. To build an ML project[5] that uses real-time data, for instance, we might collect information from several sensors to build an IoT system. The data set may come from a file, database, sensor, or a number of other sources, but it cannot immediately be used for analysis because it may have a significant amount of missing data, unusually high values, disorganised text data, or noisy data[6].
In order to solve this issue, Data Preparation has been carried out. One of the most crucial processes in machine learning is data pre-processing. It's the phase that will have the biggest impact on how accurate machine learning models become. There is a 75/25 rule in machine learning. Data pre-processing should take up 75% of a data scientist's effort, and real analysis should take up 25% of their time.

V. Machine Learning Implementation

A classification problem arises when the goal variable is categorical (i.e., the output can be divided into classes and belong to Class A, B, or something else).

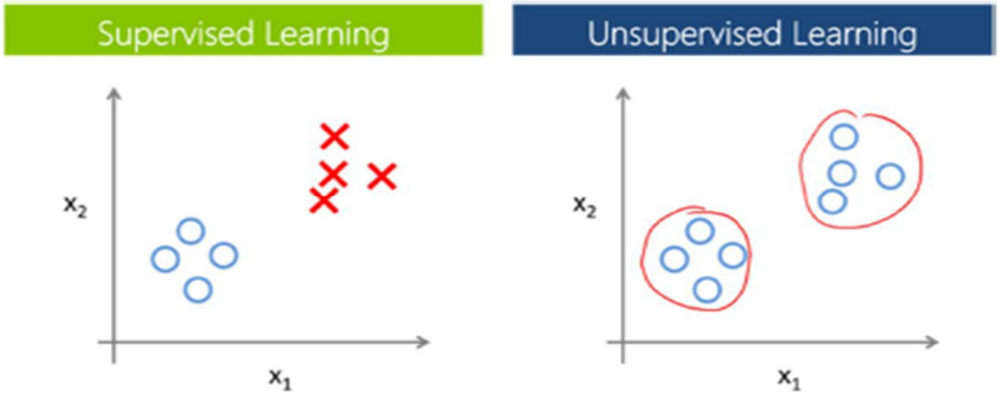

Fig. 2. Classification Techniques

VI. Algorithms for Classification

A. K-Nearest Neighbor

One of the most fundamental yet crucial categorization methods in machine learning is K-Nearest Neighbors. It falls under the category of supervised learning and has numerous applications in data mining, intrusion detection, and pattern recognition. Because to its non-parametric nature, which means that it makes no underlying assumptions about the distribution of data, it is frequently disposable in real-world circumstances (as opposed to other algorithms such as GMM, which assume a Gaussian distribution of the given data). We are provided some prior information (also known as training data), which organises coordinates according to an attribute[7].

B. Naive Bayes

The Nave Bayes algorithm is a supervised learning method for classification issues that is based on the Bayes theorem. It is mostly employed in text categorization with a large training set. The Naive Bayes Classifier is one of the most straightforward and efficient classification algorithms available today. It aids in the development of quick machine learning models capable of making accurate predictions. Being a probabilistic classifier, it makes predictions based on the likelihood that an object will occur. Spam filtration, Sentimental analysis, and article classification[8] are a few examples of Naive Bayes algorithms that are frequently used.

C. Decision Trees/Random Forest

A supervised learning method called a decision tree can be used to solve classification and regression problems, but it is typically favoured for doing so. It is a tree-structured classifier, where internal nodes stand in for a dataset's features, branches for the decision-making process, and each leaf node for the classification result. The Decision Node and Leaf Node are the two nodes of a decision tree. Whereas Leaf nodes are the results of decisions and do not have any more branches, Decision nodes are used to create decisions and have numerous branches. The given dataset's features are used to execute the test or make the decisions. It is a graphical depiction for obtaining all feasible answers to a choice or problem based on predetermined conditions. It is known as a decision tree because, like a tree, it begins with the root node and grows on subsequent branches to form a structure resembling a tree. The CART algorithm[9], which stands for Classification and Regression Tree algorithm, is used to construct a tree. A decision tree only poses a question and divides the tree into subtrees according to the response (Yes/No).

D. Support Vector Machine

A supervised machine learning approach called Support Vector Machine (SVM) is used for both classification and regression. Although we also refer to regression concerns, categorization is the most appropriate term. Finding a hyperplane in an N-dimensional space that clearly classifies the data points is the goal of the SVM method. The number of features determines the hyperplane's size. The hyperplane is essentially a line if there are just two input features. The hyperplane turns into a 2-D plane if there are three input features. Imagining something with more than three features gets challenging[10].

E. Logistic Regression

One of the most often used Machine Learning algorithms, within the category of Supervised Learning, is logistic regression. With a predetermined set of independent factors, it is used to predict the categorical dependent variable. In a categorical dependent variable, the output is predicted via logistic regression[11]. As a result, the result must be a discrete or categorical value. Rather of providing the exact values of 0 and 1, it provides the probabilistic values that fall between 0 and 1. It can be either Yes or No, 0 or 1, true or false, etc. With the exception of how they are applied, logistic regression and linear regression are very similar. Whereas logistic regression is used to solve classification difficulties, linear regression is used to solve regression problems.

To begin training a model, we divided it into three categories: "Training data," "Validation data," and "Testing data." A "training data set" is used to train the classifier, a "validation set" is used to fine-tune the parameters[12], and a "unseen test data set" is used to assess the classifier's performance. It's important to keep in mind that the classifier only has access to the training and/or validation sets during training. The test data set must not be used while the classifier is being trained. The test set will only be accessible while the classifier is being evaluated.

Fig. 3. Training and Test Dataset

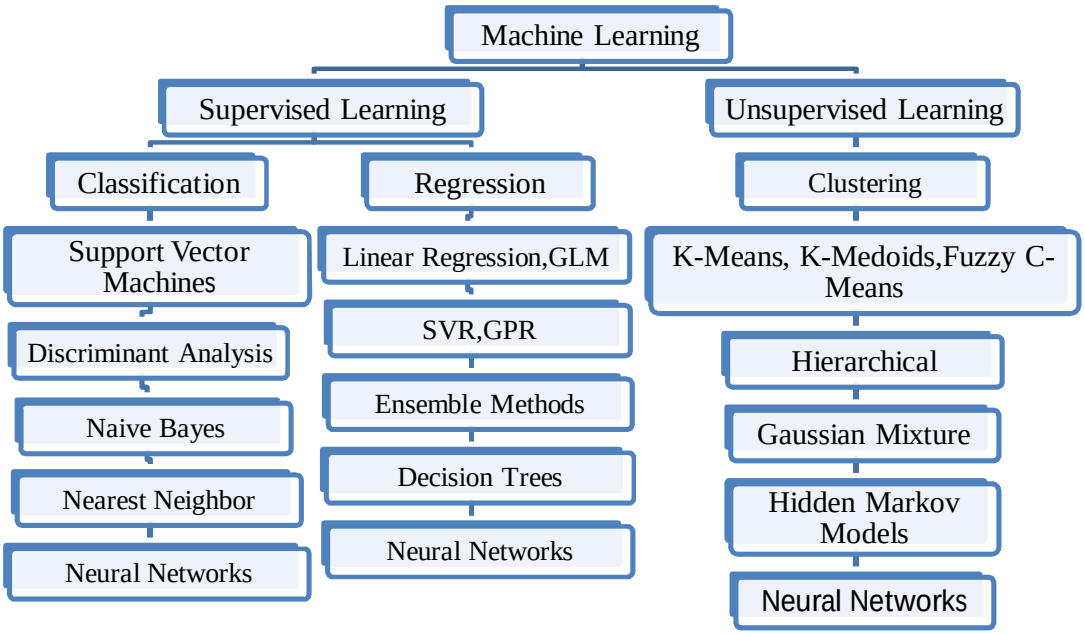

Fig. 4. Machine Learning Algorithms

The goal of the study is to suggest crops to farmers using a Decision Tree Classifier. The primary approach for this project is to pre-process the data that was provided to us, use it to build the backend model, and connect it to the Frontend interface using flask to show the whole and finished product. In contrast to the former system, which only took into account one state, we used a big dataset that covered all of India's states in our new method. These recommendations could be taken out and applied to teach the farmers.
By creating a visual representation, the farmer can better comprehend the crops to grow[13]. The dataset contains 821 distinct data points. 14 columns make up the dataset, which are explained below.
1) States: There are a total of 48 states in India.
2) Ground Water: The total amount of ground water
3) Millimetres of precipitation
4) Temperature: Celsius is the unit of measurement for temperature.
5) Soil type: There are numerous different types of soil.
6) Season: When should crops be planted?
7) Crops: A variety of crops
8) Fertilizers required: Required Fertilizer Types
9) The cost of cultivation: Overall Cost of Cultivation
10) Projected revenue: Total expected revenue
11) Seeds per hectare: the number of seeds per hectare.
12) The length of the cultivation period in days
13) Crop demand (nineteenth): Crop demand (ninth)
14) Which crops can be grown in a mixed cropping system?

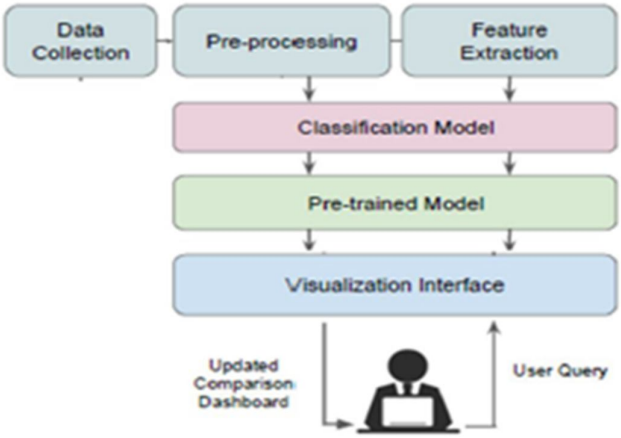

Fig 5. Machine Learning Algorithms

Information gathering and training preparation. Remove duplicates, fix errors, handle missing values, do normalisation, convert data types, and any other necessary cleaning. To mitigate the effects of the order in which we collected and/or otherwise prepared our data, randomise the data. Do additional exploratory analysis, such as data visualisation, to help find significant relationships between variables or class disparities (bias alert). Sets have been developed for training and evaluation[14].

VII. Conclusion

In-depth discussion of the significance of crop management was included in this study. Modern technology is necessary for farmers to raise their crops. Accurate crop predictions can be promptly communicated to agriculturists. Many Machine Learning techniques were used to analyse the agricultural aspects. The dataset's accessibility and the purpose of the study influence the features that are selected. Studies suggest that models with more attributes may not necessarily perform at the maximum level in yield prediction. To choose the model with the best performance, models with both more and fewer features should be compared. Several algorithms have been used in many types of study. The results suggest that some machine learning models are used more frequently than others, even though it is impossible to say for sure which model is the best from them. Including the model, we discovered an average accuracy of 95%.

REFERENCES

[1] Shreya S. Bhanose, Kalyani A. Bogawar (2016) “Crop And YieldPrediction Model”, International Journal of Advance ScientificResearch and Engineering Trends, Volume 1,Issue 1, April 2016

[2] I. Ahmad, U. Saeed, M. Fahad, A. Ullah, , A. Ahmad, J. Judge Yield forecasting of spring maize using remote sensing and crop modeling (10) (2018)

[3] Jitendra Singh Kushwah et. al. “A Comprehensive System for Detecting Profound Tiredness for Automobile Drivers Using a CNN”, Advanced Computing and Intelligent Technologies Proceedings of ICACIT 2022 Part of the book series: Lecture Notes in Electrical Engineering (LNEE, volume 914), Print ISBN: 978-981-19-2979-3, Online ISBN: 978-981-19-2980-9, 31 August 2022, Springer, Singapore.

[4] Ali, F. Cawkwell, E. Dwyer, S. GreenModeling managed grassland biomass estimation by using multitemporal remote sensing data—a machine learning approachIEEE J. 10 (7) (2017).

[5] Rabi Narayan Behera, Kajaree Das, “A Survey on Machine Learning: Concept, Algorithms, and Applications “ Vol. 5, Issue 2, February 2017.

[6] Jitendra Singh Kushwah et. al. “Analysis and visualization of proxy caching using LRU, AVL tree and BST with supervised machine learning” in 1st International Conference on Computations in Materials and Applied Engineering–2020 of “Materials Today: Proceedings” DOI: https://doi.org/10.1016/j.matpr.2021.06.224 ISSN: 2214-7853.

[7] Cheng, H.; Damerow, L.; Sun, Y.; Blanke, M. Early Yield Prediction Using Image Analysis of Apple Fruit and Tree Canopy Features with Neural Networks. J. Imaging 2017.

[8] Jitendra Singh Kushwah, Aditya Vidyarthi et. al. “COMPARATIVE ANALYSIS OF LINEAR REGRESSION, RANDOM FOREST AND SUPPORT VECTOR MACHINE USING DATASET” in International Journal for Innovative Engineering and Management Research. Vol. 11 Issue 1 pg. 30-35, ISSN: 2456-5083, Jan 2022.

[9] Gulden Kaya Uyanil, Nese Guler,” A study on multiple linear regression analysis “, Procedia - Social and Behavioral Sciences 106 (2013) 234 – 240.

[10] Jitendra Singh Kushwah et. al. “Comparative study of regressor and classifier with decision tree using modern tools” in First International Conference on Design and Materials (ICDM)-2021 of “Materials Today: Proceedings” ISSN: 2214-7853, Volume 56, Part 6, 2022, Pg 3571-3576.