

**FALL-SEMESTER**

**Course Code:** MCSE601L **Name :** Nidhi singh

**Course-Title:** – ARTIFICIAL INTELLIGENCE  **Reg. No:** 22MAI0015

**DIGITAL ASSIGNMENT - II**

**Faculty:** SIVA SANKARI S - SCOPE **Slot-**L53+L54

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Review the supervised learning algorithms in the Agriculture. (Refer minimum 5 journal papers)**

**1. Introduction**

**1.1. Importance of Application**

Agriculture can provide for one of humanity's essential needs, which is food. The globe over, agriculture is seen as a source of employment in addition to meeting the basic necessities of people. In developing nations like India, agriculture is seen as the basis of the economy and a source of jobs. In India, 15.4% of GDP is derived from agriculture. It is regarded as the main employer in the majority of the nations. Many nations, including India, still rely on traditional farming methods. Farmers there are hesitant to use cutting-edge technologies in their operations either due to ignorance of the benefits of such technologies, their high cost, or both. Lack of understanding about soil types, yields, crops, weather, inappropriate pesticide use, irrigation issues, incorrect harvesting, and ignorance of market trends contributed to the loss of farmers or increased costs. Lack of knowledge at each stage of agriculture creates new issues or worsens existing issues, which raises the cost of farming. Day by day population growth adds to the demand on the agricultural industry. Overall losses in the agricultural operations, from crop selection through product sale, are relatively significant. According to the adage "Information is Power," farmers may be able to make better decisions and resolve agricultural issues by keeping track of information on the crops, environment, and market. Pre-harvest, harvest, and post-harvest activities in agriculture are typically divided into these three categories. Machine learning research advancements have improved agricultural gains. By offering detailed advice and insights into the crops, machine learning is a modern technological advancement that helps farmers reduce their farming losses. a thorough examination of the most recent machine learning applications in agriculture to solve issues in the three stages leading up to, during, and following harvest. The use of machine learning in agriculture enables more precise and effective farming with lower human labour requirements and high-quality output. Information can be collected and processed using technologies including edge computing, cloud computing, blockchain, IoT, machine learning, and deep learning. Applications of computer vision, machine learning, and the Internet of Things will help to boost productivity, enhance quality, and ultimately increase the profitability of farmers and related industries. The improvement of the overall harvesting output depends heavily on precision learning in the agricultural sector.

Farmers typically follow the following procedures when completing agricultural operations.

Step 1: Selection of Crop

Step 2: Land Preparation

Step 3: Seed Sowing

Step 4: Irrigation & fertilizing

Step 5: Crop Maintenance [use of pesticides, crop pruning etc.]

Step 6: Harvesting

Step 7: Post-Harvesting activities

**1.2. Open Problems Associated with Application**

Agricultural issues are getting worse as the population grows. However, it is more difficult to fix issues when there are water shortages and constrained land resources. Built on the pillars of technology, modern agriculture is expanding in strength. The ability of technology to solve bigger issues is demonstrated by modern agriculture.

**1.2.1. Crop management**

## Future yields are determined by crop management, which is a significant portion of pre-harvesting tasks. The most difficult phase of the agricultural lifecycle is this one, though. Crop resistance may be impacted by an increase in the frequency of drought, rising temperatures, and unpredictable soaking and drying cycles. As a result, this stage is frequently enhanced by machine learning advancement. A technique and application of machine learning in agriculture is the selection of crop varieties, for instance. Crops should have the correct gene sequence to become disease- and weather-resistant. Crop breeding can be made easier with ML-based deep learning. The process of creating a probabilistic model is simply accomplished by algorithms gathering field data on plant behaviour.

## 1.2.2. Precision spraying

## Spraying is crucial for crop health because it helps keep diseases and pests from invading crops. This topic is also covered by agricultural machine learning projects. The agricultural industry has developed a system called precision or targeted spraying that combines the best aspects of artificial intelligence and computer vision. As a result, the system gathers target data about the plant, such as its size and shape, and then uses herbicides as necessary. The advantage of this method is that it makes it possible to apply pesticides and fertilisers more precisely according to the type of crop. To decide which chemicals should be sprayed on plants, soil, and other objects, precision spraying uses pictures and spectral signatures of those objects. By reducing crop damage risk and increasing crop output, this technology.

## 1.2.3. Insect detection

For crops in agricultural facilities, insects pose a serious hazard. Pests rob the world's crop production of between 20 and 40 percent of its annual output. Farmers employ insecticides to protect the building, but these also kill other little pests that reside nearby as well as the insects that are inside. Identifying the "bad actors" is very challenging when done manually. Drones have been used in farming for a while to spot insects, but recently there has been a rise in the use of machine learning to this task. To classify pests so that they can be captured and identified, machine learning businesses assist farmers.

**1.2.4. Right Seed –> Right Area**

A terrible issue currently affects agriculture. When an agriculture industry is suffering a great loss despite the proper knowledge being consumed. Why? Simply speaking, there is inadequate crop supervision. Finding the best location for your crop can help you reduce damage and increase revenue. Classification evaluation can also assist in this process.

**1.2.5. Field conditions management**

The management of soil and water is one of the most crucial parts of farming. However, managing all of the variables might be challenging given their sheer quantity. For this reason, farmers compile data on the soil and water conditions on their properties. The latter can then be fed into a computerised system, which will subsequently generate suggestions for fertiliser application rates, pest management strategies, and irrigation schedules based on the prediction.

**1.2.6. Yield mapping**

The goal of every farmer is to maximise the yield of their crops and to maximise the rewards. Planning your harvest and ascertaining the potential yield of each field are two ways to achieve this. The concept of "sensing and mapping" is one of the most fascinating advancements in this field. The term "trick" refers to the mapping of a field's yield utilising imaging methods and digital image processing. Yield mapping is essentially a component of precision farming. It assists in bringing out the variations in soil types throughout the property. Furthermore, mapping provides information on moisture content and enables the farmer to handle a number of connected issues on the farm.

**1.2.7. Precision spraying**

Spraying is crucial for crop health because it helps keep diseases and pests from invading crops. This topic is also covered by agricultural machine learning projects. The agricultural industry has developed a system called precision or targeted spraying that combines the best aspects of artificial intelligence and computer vision. As a result, the system gathers target data about the plant, such as its size and shape, and then uses herbicides as necessary. The advantage of this method is that it makes it possible to apply pesticides and fertilisers more precisely according to the type of crop. In order to decide which chemicals should be used, precision spraying uses pictures and spectral signatures of plants, soil, and other substances. With this method, crop output is increased while crop damage risk is reduced. A notable example of this use case is Israeli company Green eye Technology. Their AI-enabled precision spraying technique has been shown to reduce the need for herbicides by 78% and expenses by more than 50%. Farmers won't need to spend money on new equipment because the software works with any brand and size of commercial sprayer.

**1.2.8. Field conditions management**

The management of soil and water is one of the most crucial parts of farming. However, managing all of the variables might be challenging given their sheer quantity. For this reason, farmers compile data on the soil and water conditions on their properties. The latter can then be fed into a computerised system, which will subsequently generate suggestions for fertiliser application rates, pest management strategies, and irrigation schedules based on the prediction. Additionally, machine learning is used to calibrate soil sensors. The combined effects then aid in the prediction of water stress and nutritional shortages. Soil moisture affects important farming operations such as crop selection, time of tilling, and harvesting. Weather information, soil and crop variables, and other factors are typically used to predict moisture. Following that, the prediction is strengthened using empirical, regression, and machine learning techniques. With the help of this application, water resource planning may be done more data drivingly, improving yields and lowering expenses.

**1.2.9. Livestock management**

Aspects of machine learning in agriculture that are particularly important are animal welfare and livestock productivity. Numerous fields can benefit from the technology. Assessments of the wellbeing of animals, forecasting animal output, and calculations of the environmental impact of livestock operations are some of these. Because of this, farmers may better understand the health of their animals by keeping an eye on their vital signs, daily activity levels, and dietary habits. Farmers in Uganda identify infections in animals two days before they start to show symptoms. Users' smart phones or PCs, an RFID reader, and a sensor-equipped chip are all required for this technology to work. The majority of health-related factors, from eating to fertility, can be detected and tracked by the programme in this way.

**1.2.10. Automatic harvesting robots**

According to a recent study, global output of fruits and vegetables has surged by more than 50% in the last 20 years. Better techniques for collecting, sorting, and ultimately delivering products to various locations are required due to this rising production rate. The top machine learning businesses are working hard to create intelligent technologies that can automate the harvesting of agricultural goods in order to achieve this. These innovations enable the robots to identify individual fruits and vegetables based on their size, shape, or colour with greater accuracy. By reducing the amount of time spent harvesting, automated robots help farmers make more money.

**1.2.11. Irrigation**

The first and most important factor in crop production achievement is effective irrigation functionality. Improved irrigation can be achieved in the following ways thanks to machine learning algorithms:

• Preserving the ideal range of soil moisture in the root zone for plant growth.

• Little labour is required to manage the irrigation operation.

• Boost the ratio of typical vegetable yield.

• Soil moisture sensors for real-time monitoring of plant requirements.

The following machine learning irrigation systems are some examples:

• Closed loop system: When to use water and how much to use.

• Open loop system: The volume of water that will be used during each irrigation.

• Time-based system: the pre-set sum to be entered into the field.

**1.2.12. Predictive Analysis**

According to some, the main figure in the agricultural industry is a choice. Making the appropriate choice will result in happier and more prosperous circumstances. One of machine learning's greatest strengths is predictive analysis, which plays a key part in tasks like:

* Precise sowing decisions
* healthy crop production estimates
* Fertiliser advice that will significantly increase a business's value.

**1.2.13. Diagnosing Soil Defects**

Calculating risks is a key component of farming, but what if the risk could be predicted and mitigated beforehand? You may generate more revenue and save a significant amount of time by using anomaly analysis to determine the strengths and weaknesses of the soil.

**1.2.14. Production forecasting using weather condition:**

Claire Monteleoni claims, "Climate is now a data concern. Incorrect weather forecasts earlier resulted in a large number of crops being lost, costing money and time. However, as technology has advanced over the years, organisations have experienced higher, more steady growth. You can forecast production more accurately by analysing meteorological data with the use of regression analysis.

**1.2.15. Weed Detection:**

Pests, plant diseases, and post-harvest losses account for 40 to 50 percent of crop output losses in developing countries. That percentage is 20 to 25% even in the United States. You can find the current thing on the farm with the aid of image analysis. Additionally, by doing image object classification, we may locate weeds on the farm to support the establishment of healthy crops.

**1.2.16. Water Treatment**

The most crucial ingredient for plant growth is the right nutrients. Unsupervised analysis' ability to identify anomalies will enable you to choose the proper amount of minerals, resulting in crops that develop more quickly than those of your rivals and yield more harvests overall.

**1.3. General context of supervised learning in Agriculture**

In order to extract value from the growing amount of data coming from diverse sources, the digital transformation of agriculture has transformed many managerial functions into artificially intelligent systems. Machine learning, a branch of artificial intelligence, has the ability to solve a wide range of problems related to the development of knowledge-based farming systems. The goal of the current study is to shed light on the application of machine learning in agriculture by carefully reviewing the most recent scholarly literature using the keywords "machine learning" in conjunction with "crop management," "water management," "soil management," and "livestock management," and in accordance with PRISMA guidelines. Additionally, it was seen that crop management was the main focus. There were several different machine learning algorithms applied, with Artificial Neural Networks' techniques being the most effective. Aside from that, the most extensively studied crops and animals were, respectively, maize, wheat, and cattle. The use of a range of sensors, some of which are mounted on satellites and unmanned ground and aerial vehicles, has also been made as a means of obtaining trustworthy input data for data analytics. In order to increase awareness of the potential benefits of utilizing machine learning in agriculture and to contribute to more organized research on this subject, it is envisaged that this study will serve as a helpful guide for all stakeholders.

**2. Background**

**2.1. Fundamentals of supervised learning: A Brief Overview**

Nowadays, the most popular branch of machine learning is supervised learning. A typical starting point for novice machine learning practitioners is supervised learning techniques. Consequently, supervised learning will be the topic of the first of these three posts. The algorithms used in supervised machine learning are built with learning by doing in mind. Since training this kind of algorithm is similar to having a teacher oversee the entire process, the term "supervised" learning was created to describe it. Inputs and the appropriate outputs are coupled together in the training data for supervised learning algorithms. In the course of training, the algorithm looks for data patterns that correspond to the expected results. An algorithm for supervised learning will take in fresh inputs after training and decide which label to assign them based on the training data. To correctly anticipate the label for newly presented input data is the goal of supervised learning models. A supervised learning algorithm can be expressed in the following way, which is its most straightforward form:

Y=f(x)

Where Y is the anticipated result, which is established by a mapping function that gives an input value x a class. The machine learning model builds the function that links input features to a forecasted output during training. Two types of supervised learning can be distinguished: **Classification** and

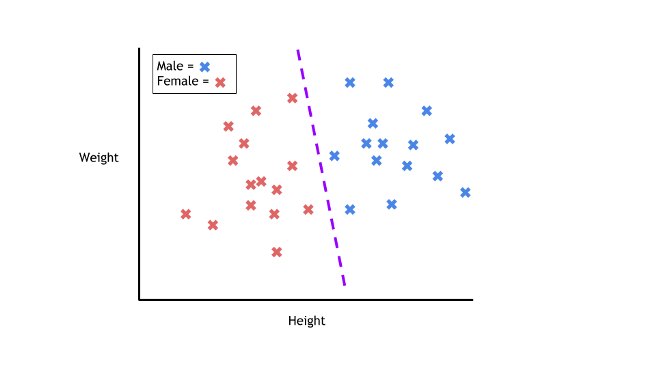
**regression**.

# ****2.1.1. Classification****

# A classification algorithm will receive data points with a designated category during training. Based on the available training data, a classification algorithm's task is to take an input value and assign it a class, or category, into which it fits.

Identifying whether or not an email is spam is the most prevalent example of classification. A binary classification problem is one where there are only two classes (spam or not spam) to pick from. Emails that are both spam and non-spam will be used as training data for the algorithm. The model will establish the aforementioned mapping function by identifying the features in the data that are associated with either class: Y=f(x). The model will then apply this function to determine if an email is spam when given an unseen email. There are several algorithms that can be used to tackle classification difficulties. The data and the circumstance will determine which algorithm you use. Several well-liked categorization algorithms are listed below:

* Linear Classifiers
* Support Vector Machines
* Decision Trees
* K-Nearest Neighbor
* Random Forest



# 2.1.2. Regression

Regression is a statistical method used to predict outcomes in which the model seeks to identify the significant link between the dependent and independent variables. To forecast a continuous number, such as sales, income, or test scores, is the aim of a regression method. The line of best fit depicts the model predictions for straightforward regression issues like this. It will be used for models with two features. The use of a hyperplane will be made for models with more than two characteristics.

Take the hypothetical situation where we wish to base a student's test grade on the number of hours they studied during the test week. The number of hours a student studies (an independent variable) and their final test grade are clearly positively correlated (dependent variable). To display the predictions made by the model given a new input, a line of best fit can be drawn between the data points. Imagine if we wanted to know how well a student would perform after five hours of study. Based on the performance of other students, we can anticipate the exam result using the line of best fit.

Regression algorithms come in a wide variety of forms. Here are the top three:

* Linear Regression
* Logistic Regression
* Polynomial Regression

Lets say the plotted data with a line of best fit looks like this:

**2.2. Brief description of the categories of problems in the application**

**2.2.1. Crop Management**

Crop management is a broad area that includes a variety of elements that result from the blending of farming methods in the direction of controlling the biological, chemical, and physical crop environment with the goal of achieving both quantitative and qualitative objectives. Utilizing cutting-edge methods for crop management, such as yield forecasting, disease and weed detection, crop identification, and crop quality, helps to boost production and, as a result, monetary income. Key objectives of precision agriculture are the aforementioned components.

**Yield Prediction**

One of the most crucial and difficult subjects in modern agriculture is yield prediction in general. For example, a precise model can assist farm owners in making knowledgeable management decisions on what to cultivate in order to match the crop to the needs of the current market. Although it only requires a few steps, this is not a simple operation. Several variables, including the environment, crop management techniques, genotypic and phenotypic traits, and their interactions, might affect yield prediction. The relationship between these interaction parameters and yield must therefore be fundamentally understood. A thorough dataset and potent algorithms, like ML methods, are required for finding these kinds of correlations.

**Disease Detection**

Crop diseases are a significant risk to agricultural production systems because they reduce output quality and quantity at the production, storage, and transportation levels. Plant disease-related yield losses are frequently reported at the farm level. Furthermore, crop diseases significantly threaten global food security. The key to effective management is the early detection of plant diseases. Different bacterial, fungal, pest, viral, and other types of agents can cause plant diseases. Leaf and fruit spots, wilting and colour changes, curving of leaves, and other indications of disease, or the physical manifestations of pathogens and changes in the phenotypic of the plants, may be present. Expert agronomists used field scouting techniques in the past to find diseases. The only basis for this method is visual assessment, which takes time. Sensing devices that are now on the market may now detect unhealthy plants before the symptoms appear thanks to recent technological advancements. Furthermore, computer vision has advanced significantly over the past few years, particularly when deep learning has been used. It is advantageous to remove background information prior to model training because of the complicated environmental background, as Zhang et al. who focused on using deep learning to identify cucumber leaf diseases have noted. Furthermore, a sizable collection of both healthy and ill plant photos is required for accurate image classifiers for disease identification. Such automated procedures can be paired with autonomous vehicles in the context of large-scale cultivations to quickly detect phytopathological issues through the use of routine inspections. Using large quantities of herbicides, however, turns out to be both costly and detrimental for the environment, especially in the case of uniform application without taking into account the spatial distribution of the weeds. Maps of the geographical distribution of the plant disease can also be produced, showing the areas of the farm where the illness has spread. Finding weeds Weeds typically develop and spread invasively across significant areas of the field very quickly because of their prolific seed production and lifespan, competing with crops for resources like space, sunlight, nutrients, and water availability. A further factor that harms crop growth is the fact that weeds usually emerge earlier than crops since they do not have to contend with natural enemies. Weed control, either through mechanical treatment or the administration of herbicides, is a crucial management duty in order to avoid crop production decline. Herbicide application is necessitated by the fact that mechanical treatment is frequently challenging to carry out and ineffectual if done incorrectly. Interestingly, repeated use of herbicides will almost certainly increase weed resistance, making weed management more difficult and expensive. On the basis of clever agriculture, significant advancements have been made recently in the separation of weeds from crops. With sensors mounted on satellites, aircraft, ground vehicles, and unmanned (both ground (UGV) and aerial (UAV)) vehicles, remote or proximal sensing can be used to achieve this differentiation. However, given the laborious nature of data gathering and classification, turning the data collected by UAVs into useful information is still a difficult undertaking. In order to accurately apply herbicides to specified zones rather than spraying entire fields and designing the shortest weeding path, ML algorithms in combination with imaging technology or non-imaging spectroscopy can provide real-time distinction and localisation of target weeds.

**Crop Recognition**

Numerous scientific disciplines, including plant taxonomy, botanical gardens, and the discovery of new species, have paid significant attention to the automatic recognition of crops. Analyzing various organs, such as leaves, stems, fruits, flowers, roots, and seeds, it is possible to identify and categorise different plant species. The most popular method of identifying plants appears to be leaf-based plant recognition, which looks at the colour, shape, and texture of individual leaves. Crop classification by remote sensing has gained a lot of traction with the expansion of the use of satellites and other aerial vehicles to detect crop attributes. Similar to the aforementioned subcategories, the automatic recognition and classification of crops has been made possible by the development of computer software and image processing hardware in conjunction with machine learning.

**Crop Quality**

In general, soil and climate conditions, cultivation techniques, and crop features, to name a few, are all related to crop quality, which has a significant impact on the market. Since better-quality agricultural products normally sell for more money, farmers can make more money. For example, among the most common maturity indices used for harvesting in terms of fruit quality are skin colour, soluble solids concentration, and flesh hardness. In both high value crops (tree crops, grapes, vegetables, herbs, etc.) and arable crops, the timing of harvesting has a significant impact on the quality features of the harvested items. Therefore, creating decision support systems can help farmers make the right management decisions for improved production quality. A management strategy like selective harvesting, for instance, may greatly improve quality. Furthermore, crop quality is intimately related to food waste, which presents another difficulty for modern agriculture to overcome because crops that don't conform to the ideal shape colour, or size may be discarded. Similar to the aforementioned part, ML algorithms in conjunction with image technology can yield promising results.

**2.2.2. Water Management**

Given that plant development heavily depends on the availability of water, the agricultural industry is the primary global consumer of fresh water. More efficient water management is required in order to better save water in order to achieve a sustainable crop output due to the rapid depletion rate of many aquifers and minimal recharge. The quality of the water itself can be improved, and pollution and health concerns can be decreased as a result of efficient water management. The potential of variable rate irrigation to achieve water savings has recently been explored in precision agriculture research. Instead of applying irrigation at a consistent rate throughout the entire field, this can be achieved by delivering irrigation at rates that vary according to field variability and the unique water requirements of different management zones. To achieve both water savings and yield optimization, the variable rate irrigation approach's performance and viability depend on agronomic parameters, such as topography, soil characteristics, and their impact on soil water. Planning irrigation and effective water management can be aided by carefully monitoring the state of the soil's water supply, crop growth conditions, and temporal and spatial trends. For precise water management, remote sensing is one of the ICTs that is used to produce images with spatial and temporal variability related to the crop growth parameters and soil moisture status. The utilisation of groundwater sources for irrigation in dry regions, where precipitation only meets a portion of the crop's total evaporate transpiration (ET) needs, makes managing water there an interesting challenge.

**2.2.3. Soil Management**

A diverse natural resource, soil, includes extremely complicated mechanisms and processes. Achieving improved soil management that is consistent with the potential of the land and, generally, sustainable agriculture, depends on precise information about soil at the regional level. Due to challenges like soil erosion, soil nutrient imbalance brought on by excessive fertiliser use, and land degradation (loss of biological productivity), better soil management is also very important (as a result of vegetation overcutting, improper crop rotations rather than balanced ones, livestock overgrazing, and unsustainable fallow periods). Soil is a variety of natural resources with incredibly complex mechanisms and processes. Accurate knowledge about soil at the regional level is necessary to improve soil management that is in line with the potential of the land and, generally, sustainable agriculture. Better soil management is essential due to issues including soil erosion, soil nutrient imbalance brought on by excessive fertiliser use, and land degradation (loss of biological productivity) (as a result of vegetation overcutting, improper crop rotations rather than balanced ones, livestock overgrazing, and unsustainable fallow periods).

**2.2.4. Livestock Management**

A range of natural resources, soil has very intricate mechanisms and processes. To improve soil management that is in line with the potential of the land and, generally, sustainable agriculture, accurate knowledge about soil at the regional level is required. Due to problems like soil erosion, soil nutrient imbalance brought on by excessive fertiliser use, and land degradation (loss of biological productivity), better soil management is crucial (as a result of vegetation overcutting, improper crop rotations rather than balanced ones, livestock overgrazing, and unsustainable fallow periods). In addition to monitoring the products' quality and animal welfare as needed by policy makers, it can also make the products' tractability easier. Non-invasive sensors, including cameras, accelerometers, gyroscopes, radio-frequency identification systems, pedometers, and optical and temperature sensors, are essential in precision cattle husbandry. IoT sensors use variable physical quantities (VPQs) to detect things like humidity, sound, and temperature. For instance, real-time alerts from IoT sensors can provide important information on specific animals if a VPQ exceeds normal bounds. As a result, it will be less expensive to check each animal one by one repeatedly and laboriously. Modern cattle husbandry now uses ML approaches on a regular basis to benefit from the abundance of data. It is possible to create models that can describe how a biological system behaves by depending on causal linkages and using this biological awareness to produce predictions and recommendations.

**Animal Welfare**

Since animal health is closely linked to product quality, which in turn has a major impact on consumer health and, indirectly, on increasing economic efficiency, there is a persistent concern for animal welfare. Numerous indications, including as behavioural and physiological stress markers, are available to assess the welfare of animals. The most widely utilised indicator is animal behaviour, which is susceptible to being influenced by ailments, feelings, and living circumstances, all of which have the potential to reveal physiological abnormalities. Mick systems, cameras, accelerometers, and other sensors are frequently used to identify behavioural changes (such as variations in water or food consumption, decreased animal activity, etc.).

**Livestock Production**

The efficiency of animal production can be raised by utilizing sensor technologies and cutting-edge ML approaches. Due to the effects of animal management practices on aspects that are productive, livestock owners are becoming more protective of their asset. The correct assessment of every single animal, however, becomes increasingly challenging as livestock holdings grow. From this vantage point, the above mentioned precision livestock farming programmed that supports farmers is a positive development for factors related to economic effectiveness and the creation of environmentally friendly workplaces. In general, a variety of models have been applied to the production of animals, with the main focus typically being on the best methods for raising and feeding animals. But once more, the substantial amounts of data involved necessitate ML strategies.

**3. Methods**

Food, one of humanity's basic requirements, is provided through agriculture, which is regarded as a major economic pillar worldwide. In most nations, it is regarded as the primary source of employment. Many countries, including India, continue to engage in conventional farming. Because of their lack of comprehension, the expensive expense, or a lack of knowledge about the advantages of such technology, farmers in that region are reluctant to adopt it. Landowners were lost or costs rose as a result of a lack of knowledge about different soil types, productivity, harvests, climate, inadequate pesticide usage, irrigation problems, incorrect harvesting, and market trends. Each stage of the agribusiness is affected by inadequate information, which boosts the cost of farming by causing new problems to arise or makes current ones worse. The agricultural sector faces increased demand as a result of daily population expansion. The entire agricultural process, from crop selection through product sale, results in large overall losses. According to the proverb "Intelligence is Power," farmers may be able to address agricultural problems and make better judgements by making notes about the crops, environment, and market. Technologies like blockchain, IoT, machine learning, deep learning, cloud computing, and edge computing can be used to gather and process data. Let's examine a few studies that summarise the growth of machine learning in the agricultural sector. How effectively the ML model performs in a certain task is determined by a performance indicator that improves over time with training. To evaluate how well ML models and procedures work, many statistical and mathematical techniques are applied. Following the conclusion of the learning process, the trained model can be used to classify, predict, or cluster new instances (testing data) in accordance with the information learned.

**Paper 1:-**

[1]Performances of Machine Learning Algorithms in Predicting the Productivity of Conservation Agriculture at a Global Scale

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **S.No** | **Author & Referen-ces** | **Objective** | **Proposed Methodology** | **How novelty introduced- in propos-ed methodolo-gy** | **Compared Methodology** | **Perform-ance metrics used** | **Remarks (output, future enhancem-ent)** |
| 1 | Yang Su et al.  (2022)  [1] | For the purpose of measuring the productivity of CA systems and analysing uncertainty in model outputs, evaluate and compare various machine-learning models. | 1.Dataset Establishme-nt  2.Model Training  random forest, random forest with spatial  correlations, ANNs, KNN, SVM, naïve bayes  3. Model Tuning With 10-Fold  Cross-  Validation  4. Global Projection | The implementation and comparison of 12 ML algorithms in an application addressing a significant, worldwide problem for agriculture is a first. | We show that  random forest has the best performance in classification and regression, while quantile  regression forest performs better than quantile neural networks in quantile regression. We consider different usages, including classification, point regression and quantile | AUC value,  Error Score,  Interval scores (IS), root-mean-square error (RMSE) | The more conventi-onal algorithms GLM and NB exhibit poorer performance, with AUC values of 0.644 and 0.647, respectively, whilst Random forest (RF), GBM, and XGBOOST show superior classificat-ion performa-nce, with AUC values equal to 0.790, 0.786, and 0.783, respective-ly. |

[1] In this study, we trained a large variety of machine learning (ML) models for categorization, quantitative prediction, and range prediction, among other uses. Our goal is to assess and contrast several machine-learning models' propensity to predict the output of CA systems as well as to examine the level of uncertainty in the model results. Classification, point regression, and quartile regression are just a few of the applications we take into account. Our methodology compares 12 alternative machine learning algorithms and includes model training, adjustment with cross-validation, testing, and results projection on a global scale. A recent worldwide dataset with more than 4,000 pairs of crop production data for CA vs. CT is used to compare the performances of various algorithms. In both classification and regression, random forest performs best, whereas quartile regression forest outperforms quartile neural networks in quantise regression. The findings of mapping crop productivity of CA vs. CT using the best methods show that the performance of CA vs. CT is characterised by a substantial spatial variability and that the likelihood of yield gain with CA is extremely location-dependent. This finding shows how much more helpful our method is than merely summarising the average impact sizes from conventional meta-analyses, and it opens the door for similar probabilistic, spatially explicit methods in many other areas of study. When evaluating and comparing quintile regression models, it provides a more precise and thorough evaluation statistic than coverage probability. While the more conventional algorithms GLM and NB perform worse than the more recent Random Forest (RF), GBM, and XGBOOST, with AUC values of 0.790, 0.786, and 0.783, respectively, the more recent Random Forest (RF), GBM, and XGBOOST all exhibit higher classification performance. Two hidden layers and an AUC value of 0.752 were used with ANN to achieve the best classification accuracy.

**Paper 2:- [2] Machine Learning in Agriculture: A Comprehensive Updated Review**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **S.No** | **Author & References** | **Objective** | **Proposed Methodology** | **How novelty introduced in proposed methodology** | **Compared Methodology** | **Perform-ance metrics used** | **Remarks (output, future enhancem-ent)** |
| 2 | Benos, L. et al.  (2021)[2] | Aims to provide a complete assessment of the most recent academic literature using the keywords "machine learning" coupled with "crop management," "water management," "soil management," and "livestock management" in accordance with PRISMA principles. | CNNs are used with the goal of teaching a classifier to distinguish between photos of healthy leaves. Input data from multispectral or hyperspectral measurements made by spectroradiometers, UAVs, and satellites were the second most popular sources. | It is envisaged that this study will serve as a useful manual for all interested parties in raising awareness of the potential benefits of machine learning in agriculture and advancing further organised research on the subject. | Artificial Neural Network,  Bayesian Models,  Deep Learning,  Dimensional-ity Reduction,  Decision Trees,  Ensemble Learning  Instance Based Models,  Support Vector Machine | F-scores, error, MAE, RMSE, relative prediction error, Accuracy | . RF: Acc = 76.3–96.5%,Higher correlation: QRM: R2 = 0.922, SVM: carcass weight: R = 0.945, MAE = 0.139,  EMA: R = 0.676, MAE = 4.793, MS: R = 0.631 |

[2] By carefully evaluating contemporary scholarly literature using the keywords "machine learning" coupled with "crop management," "water management," "soil management," and "livestock management," and following PRISMA principles, the current study intends to shed light on machine learning in agriculture. According to the findings, this subject is relevant to various fields that support global convergence research. Additionally, it was seen that crop management was the main focus. There were many different machine learning algorithms utilised, with artificial neural network based techniques being the most effective. Additionally, the most extensively studied crops and animals were wheat and maize, along with cattle and sheep. In order to obtain trustworthy input data for the data analytics, a range of sensors mounted on satellites and unmanned ground and aerial vehicles have been used. It is envisaged that this study will serve as a useful manual for all interested parties in raising awareness of the potential benefits of machine learning in agriculture and advancing further organised research on the subject.

**Paper 3:-** [3] Smart Farming Prediction Using Machine Learning

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **S.No** | **Author & References** | **Objective** | **Proposed Methodology** | **How novelty introduced in propos-ed method-ology** | **Compared Methodology** | **Performa-nce metrics used** | **Remarks (output, future enhancem-ent)** |
| 3 | S.R.Rajeswari et al. (2019)[3] | The methods used to forecast agricultural yield and crop costs are the main topics of this research. All of these traits can be used to implement smart farming. | A given dataset's attribute is statistically analysed using a Bayesian network. Following that, the ANN is used to compare patterns with nonlinear effects and highlight the concept of the relationship between them; as a result, it is a type of ML technique with a large amount of memory. | The proposed system is a depiction of all the parts and aspects that, upon first glance, should provide a person with a general overview and also help them to understand the project's skeleton. | Random forest algorithm, Clustering, Bayesian Network, Artificial Neural Network, | After applying the classification techniques we get two results, Algorithm result (accuracy of the datasets) ,Dataset results that will be in the form of a matrix such as true positive ,true negative etc.). | The system will be fed with 3000 general data of agricultural aspects using datasets from koggle.com as part of the implementation. Among them include temperature, soil quality, etc. Two types of data -trained data and test data - are needed for the machine learning algorithms to use the prediction system. |

[3] The system will be fed with 3000 general data of agricultural aspects using datasets from koggle.com as part of the implementation. These comprise temperature, soil condition, etc. Two types of data—trained data and test data—are needed for the machine learning algorithms to use the prediction system. Proposed System is a depiction of all the parts and pieces that, at a glance, should provide a person with a basic overview and help them understand the framework of the project. The Davis Pro2 is a piece of hardware that allows for the regular collection of field data utilising sensors and transmission to cloud storage. All field data will be contained in cloud storage, which will also track any data modifications. As needed, all of these data will be transformed into datasets. As the method is implemented here, the classification technique is the most crucial step in the process. The Random Forest method is used in the process to provide comprehensive findings for the datasets. The algorithm uses the remaining 80% of the train data and 20% of the test data (random data) as input. We receive two results after using classification techniques: an algorithm result (dataset accuracy) and dataset results (which take the form of a matrix and include true positive, true negative, and other categories). From the matrix itself, one can infer the predicted data. To make a prediction of the land to grow the desired crop in a specific feature of the month, the values of the matrix are employed in real time. Approximately 20% of the test data and 80% of the training data will each undergo random processing as part of the pre-processing of these data. The Random Forest algorithm will be used to forecast the percentage of the crop that will grow in the kind of soil over the course of several months. Research will be completed once the projected data has been obtained to see whether or not it is accurate in real time. Crop will be evaluated using a matrix format, with true-positive and true-negative values.

**Paper 4:-** [4] Machine Learning in Agriculture Application: Algorithms and Techniques

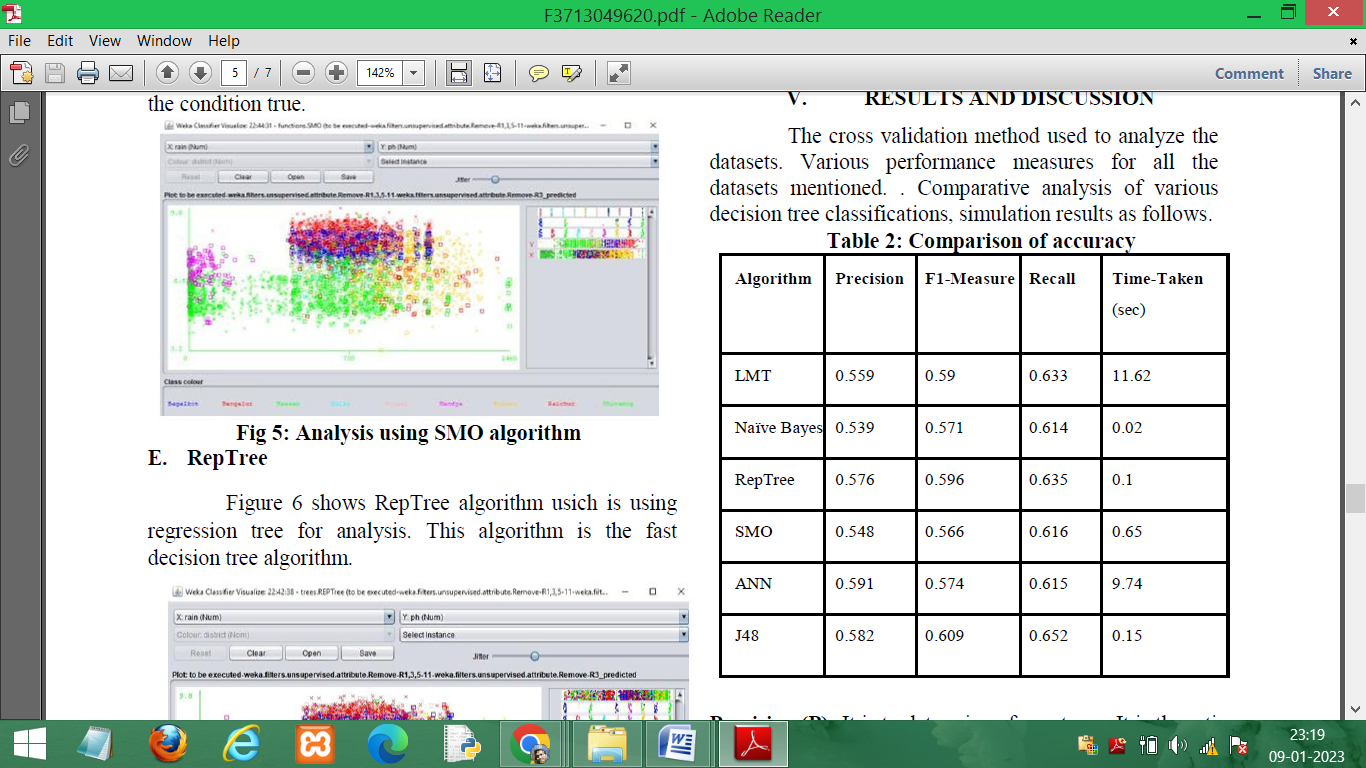
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **S.No** | **Author & References** | **Objective** | **Proposed Methodology** | **How novelty introduced in proposed methodology** | **Compared Methodology** | **Perform-ance metrics used** | **Remarks (output, future enhancem-ent)** |
| 4 | Meeradevi,et al. (2020)[4] | The results of this survey report show that ML techniques outperform numerous traditional techniques in prediction and Deep learning models provide excellent accuracy and outperform standard image processing techniques. | A data collection encompassing various crops, rainfall, production, area, and pH variables was subjected to the categorization data mining methods. The information includes soil parameters like pH and rainfall. | This paper conducts a thorough assessment of numerous papers that focus on the use of machine learning (ML) and deep learning in agriculture. | LMT  Naïve Bayes  RepTree  SMO  ANN  J48 | Precision,  F1-Measure,  Recall,  Time-Taken | The datasets were examined using the cross validation approach. Naive bayes is the fastest, processing in 11.62 seconds, while LMT is the slowest. |

[4] A wide range of new opportunities in the agriculture sector can be created by combining machine learning techniques with high performance computer technologies. The application of machine learning (ML) and deep learning to agriculture is the focus of this research, which conducts an extensive evaluation of numerous papers in the field. The three components of this essay are as follows:

a) Yield prediction using machine learning technique

b) Price prediction

c) Leaf disease detection using neural networks.

In this article, we investigate how neural network models compare to already-existing models. The results of this survey report show that ML techniques outperform numerous traditional techniques in prediction and Deep learning models provide excellent accuracy and outperform standard image processing techniques. The datasets were examined using the cross validation approach. Different performance metrics for each of the aforementioned datasets. The following are the simulation findings after a comparative investigation of different decision tree classifications. ******

**Paper 5:-** [5] Comparative Analysis of Machine Learning Algorithms in The Study of Crop and Crop yield Prediction

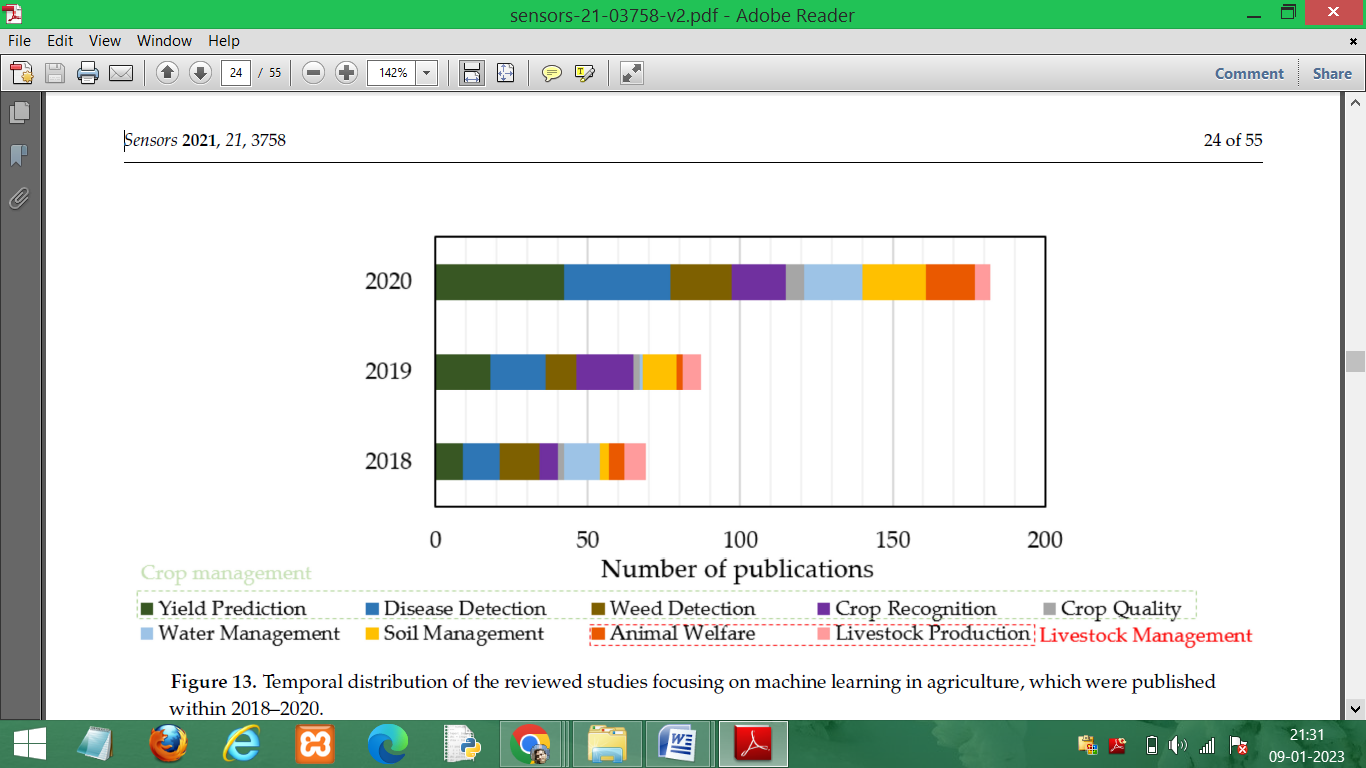
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **S. No** | **Author & References** | **Objective** | **Proposed Methodology** | **How novelty introduced in proposed methodolog-y** | **Compared Methodology** | **Perfor-mance metrics used** | **Remarks (output, future enhancement)** |
| 5. | S Bharath. et al. (2020)[5] | The framework includes a model that is precise and excellent at predicting crop yield and provides the end user with adequate information about the harvest that may depend on the climatic and soil conditions of the land, both of which are improvements that increase harvest yield and increase rancher turnover. | Based on input parameters for rainfall, soil pH, and temperature, crop and crop production predictions are made. The paper also displays a comparative analysis of algorithms with respect to the degree of accuracy and r2 score they attained and makes recommendations for the best algorithm that may be used to estimate the crop and yield, respectively. | A number of regression algorithms are used to estimate the yield. | Support Vector Machine,  Decision Tree Classifier,  k-NN Classifier and  Random Forest Classifier , Linear Regression,  Decision Tree Regressor,  Random Forest Regressor, ElasticNetCV | Accuracy, Root Mean  Square Error score as negative and r2 score | Support Vector Machine -92.6%,  Decision Tree Classifier- 99.87%,  k-NN Classifier- 99.73%,  Random Forest Classifier -81.07% |

[5] According to agricultural scientists, the development of the national economy depends on an efficient system playing a crucial role. Agribusiness makes up around 20% of the GDP in a country like India. According to the climate measuring factors, such as temperature, precipitation, and moreover soil ph, this expectation will help the ranchers choose appropriate harvests for development in their homestead. All of these authorised pieces of information will be scrutinised. For the purpose of creating the required model, we will prepare the data using several suitable ML calculations. The framework includes a model that is accurate and precise in predicting crop production and provides the end user with adequate information about the harvest, which can be influenced by the climatic and soil conditions of the land, which can increase harvest yield and increase turnover. In order to anticipate crops and crop yields, this research shows how several ML algorithms may be used effectively. According on input parameters for rainfall, soil pH, and temperature, crop and crop yield are predicted. A comparative analysis of algorithms is also shown in the paper, along with recommendations for the best algorithm to use in predicting the crop and yield, respectively. Implementing a user-friendly interface with multiple language options and further developing the system with IOT devices are the next tasks.

**4. Conclusion**

The current systematic review study focuses on ML in agriculture, a subject that is getting more and more attention globally. In order to do this, a thorough investigation of the current situation with regard to the four broad categories that Liakos et al. had defined in their earlier evaluation was conducted. These subcategories deal with managing land, water, crops, and livestock. As a result, numerous elements were examined using an integrated methodology by analysing the pertinent literature from the previous three years (2018–2020). In summary, it is possible to infer the following key conclusions:

* Crop management was the main topic of most journal papers, with the other three generic groups contributing nearly equally. Using the review paper as a benchmark, it can be inferred that the aforementioned picture is essentially unchanged, with the percentage of publications mentioning livestock declining from 19% to 12% in favour of those mentioning crop management. But this only illuminates one side of the story. Approximately 400% more publications on cattle management were discovered when the number of related papers published during the recent three years was taken into consideration (in particular, 40 articles were found in comparison to the 338 of the present literature review). The growing study focus on crop recognition was a significant additional discovery.
* Several ML algorithms have been developed to handle the heterogeneous data coming from agricultural fields. Families of ML models can be used to group these methods. ANNs were shown to be similarly effective as ML models. However, interest has also changed to EL, which may aggregate predictions from multiple models, in contrast to this. SVM completes the group of the three most accurate ML models in agriculture because to several advantages, such as its great performance while working with picture data.
* In terms of the crops that have been the subject of the greatest research, maize has received the most attention, followed by research on wheat, rice, and soybean. Cattle, together with sheep and goats, stood out among livestock management studies, making up over 85% of them. While wheat, rice, and cattle continue to be crucial specimens for ML applications, more species have been added as compared to.
* The display of the input data used in the ML algorithms and the accompanying sensors was a crucial outcome of the current review study. The most popular option was RGB photos, which justifies the widespread use of CNNs because they can process this kind of data more well. Additionally, a wide range of weather-related data as well as soil, water, and crop quality parameters were used. Remote sensing, which includes images from satellites, UAVs, and UGVs, as well as in-person and laboratory tests, was the most popular method of gathering measurements for ML applications. As was already mentioned, UAVs continue to outperform satellites in part due to their adaptability and capacity to deliver high-resolution photographs in all weather conditions. On the other hand, time-series can be provided over vast areas by satellites. The research pertaining to livestock production examined the major physical and development characteristics of the animal, whereas those pertaining to animal wellbeing used primarily sensors like accelerometers for activity recognition. Another interesting result of the current investigation is the capture of the growing interest in subjects involving ML analyses in agricultural applications. If a comparison is made between 2018 and 2019, more specifically, a 26% rise in the overall number of the pertinent research was given. The following year (2020), the equivalent increase soared to 109% compared to the results of 2019, culminating in a total increase of 164% when compared to 2018. Following the significant developments of ICT systems in agriculture, a number of reasons have contributed to the increasing rate of research interest in ML. Additionally, it is imperative to improve agricultural methods' effectiveness while alleviating their impact on the environment. In order to provide a comprehensive perspective of the processes occurring in agriculture, this necessitates the processing of both huge volumes of data and trustworthy measurements. The current technology upheaval has a big potential to support agriculture and move it in the direction of boosting food security and meeting the expanding consumer demands.



Overall, with artificial intelligence in agriculture becoming more and more valued, machine learning (ML) will undoubtedly play a supporting role in the development of a sustainable and productive farm. The current systematic effort is predicted to serve as a helpful manual for researchers, manufacturers, engineers, ICT system developers, policymakers, and farmers and, as a result, contribute to a more systematic study of ML in agriculture.

**5. References**

[1] Su Y, Zhang H, Gabrielle B and Makowski D, 2022. Performances of Machine Learning Algorithms in Predicting the Productivity of Conservation Agriculture at a Global Scale. Front. Environ. Sci. 10:812648. doi: 10.3389/fenvs.2022.812648.

[2] Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. Machine Learning in Agriculture: A Comprehensive Updated Review. Sensors 2021, 21, 3758. <https://doi.org/10.3390/> s21113758.

[3]S.R.Rajeswari , Parth Khunteta, Subham Kumar,Amrit Raj Singh,Vaibhav Pandey, May, 2019 , Smart Farming Prediction Using Machine Learning, (IJITEE) ISSN: 2278-3075, Volume-8 Issue-7.

[4] Meeradevi, Sindhu N, Monica R Mundada, , April 2020. Machine Learning in Agriculture Application: Algorithms and Techniques, (IJITEE) ISSN: 2278-3075 (Online), Volume-9 Issue-6.

[5] S Bharath, Yeshwanth S, Yashas B L,Vidyaranya R Javalagi, 2020.Comparative Analysis of Machine Learning Algorithms in The Study of Crop and Crop yield Prediction, Published by (IJERT), ISSN: 2278-0181,Volume 8, Issue 14, www.ijert.org.