Machine Learning In Agriculture

Done By

Bojjireddy Raghunadh

INTRODUCTION

Due to the population growth day by day food requirement will increase to 70% at the end of 2050 and as we can see also due to the rapid urbanization land availability for agriculture will decrease drastically in the coming years.so because of lack of planning and unpredictable weather condition,improper harvesting and irrigation technique we doesn’t meet the food production requirement.here the concept of smart farming emerge that make agriculture more efficient and effective with the help of high-precision algorithms.with the help of new technologies and high-performance computing we can understand data intensive processes in agricultural operational environments.

Machine learning is everywhere throughout the whole growing and harvesting cycle. It begins with a seed being planted in the soil—from the soil preparation, seeds breeding and water feed measurement—and it ends when robots pick up the harvest determining the ripeness with the help of computer vision

How agriculture can benefits from machine learning

- **Species breeding**-Machine learning, in particular, deep learning algorithms, take decades of field data to analyze crops performance in various climates and new characteristics developed in the process. Based on this data they can build a probability model that would predict which genes will most likely contribute a beneficial trait to a plant.

- **Species recognition**-While the traditional human approach for plant classification would be to compare color and shape of leaves, machine learning can provide more accurate and faster results analyzing the leaf vein morphology which carries more information about the leaf properties

Crop management

Yield prediction-

Yield prediction is one of the most important and popular topics in precision agriculture as it defines yield mapping and estimation, matching of crop supply with demand, and crop management. Computer vision technologies to provide data on the go and comprehensive multidimensional analysis of crops, weather, and economic conditions to make the most of the yield for farmers and population by maintaining the desired soil water range in the root zone sensors that collect soil moisture data for monitoring plant needs at real-time all these new technique will result in increase the average vegetable yield.

Crop Quality-

The accurate detection and classification of crop quality characteristics can increase product price and reduce waste. In comparison with the human experts, machines can make use of seemingly meaningless data and interconnections to reveal new qualities playing role in the overall quality of the crops and to detect them

PROBLEM STATEMENT

Crop prediction based on Weather Conditions

THE DATASET

- Feature consist of N, P, K, Temperature, Rainfall, humidity, ph and label for the type of crop.
  N - ratio of Nitrogen content in soil - kg/ha
  P - ratio of Phosphorus content in soil - kg/ha
  K - ratio of Potassium content in soil - kg/ha
  temperature - temperature in degree Celsius
  humidity - relative humidity in %
  ph - ph value of the soil
  rainfall - rainfall in mm
- Number of dataset =2200, train set =1774, test set =440
- There is no NULL value for any attribute.
- Labels =22
- For each labels we have 100 instances of data

Link for dataset : https://www.kaggle.com/mohsinisu/crop-recommendation-tidymodels/data

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

MACHINE LEARNING METHODS

1. Decision Tree Classifier
2. Random Forest
3. K-Nearest Neighbours
4. Gaussian Naive Bayes
5. Support Vector Machine

1. Decision Tree

- Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems.
- It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.
- Accuracy of the model: 0.9818181818181818

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	23
banana	1.00	1.00	1.00	21
blackgram	0.95	1.00	0.98	20
chickpea	1.00	1.00	1.00	26
coconut	1.00	1.00	1.00	27
coffee	1.00	1.00	1.00	17
cotton	1.00	1.00	1.00	17
grapes	1.00	1.00	1.00	14
jute	0.92	0.96	0.94	23
kidneybeans	1.00	1.00	1.00	20
lentil	0.92	1.00	0.96	11
maize	1.00	0.95	0.98	21
mango	1.00	1.00	1.00	19
mothbeans	1.00	0.92	0.96	24
mungbean	1.00	1.00	1.00	19
muskmelon	1.00	1.00	1.00	17
orange	0.93	1.00	0.97	14
papaya	1.00	1.00	1.00	23
pigeonpeas	1.00	1.00	1.00	23
pomegranate	1.00	1.00	1.00	23
rice	0.94	0.89	0.92	19
watermelon	1.00	1.00	1.00	19
accuracy			0.99	440
macro avg	0.98	0.99	0.99	440
weighted avg	0.99	0.99	0.99	440

2. Support Vector Machine

- SVM model is used both for Classification and Regression Models.

- SVM tries to find the “best” margin (distance between the line and the support vectors) that separates the classes and thus reduces the risk of error on the data.
- Accuracy for this model : 0.9772727272727273

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	23
banana	1.00	1.00	1.00	21
blackgram	0.95	1.00	0.98	20
chickpea	1.00	1.00	1.00	26
coconut	1.00	1.00	1.00	27
coffee	1.00	1.00	1.00	17
cotton	0.94	1.00	0.97	17
grapes	1.00	1.00	1.00	14
jute	0.85	0.96	0.90	23
kidneybeans	0.91	1.00	0.95	20
lentil	0.92	1.00	0.96	11
maize	1.00	0.95	0.98	21
mango	1.00	1.00	1.00	19
mothbeans	1.00	0.96	0.98	24
mungbean	1.00	1.00	1.00	19
muskmelon	1.00	1.00	1.00	17
orange	1.00	1.00	1.00	14
papaya	1.00	1.00	1.00	23
pigeonpeas	1.00	0.87	0.93	23
pomegranate	1.00	1.00	1.00	23
rice	0.94	0.79	0.86	19
watermelon	1.00	1.00	1.00	19
accuracy			0.98	440
macro avg	0.98	0.98	0.98	440
weighted avg	0.98	0.98	0.98	440

3. Gaussian Naive Bayes

- Naive Bayes is a statistical classification technique based on Bayes Theorem. It is one of the simplest supervised learning algorithms.

- Naive Bayes classifier assumes that the effect of a particular feature in a class is independent of other features.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	23
banana	1.00	1.00	1.00	21
blackgram	1.00	1.00	1.00	20
chickpea	1.00	1.00	1.00	26
coconut	1.00	1.00	1.00	27
coffee	1.00	1.00	1.00	17
cotton	1.00	1.00	1.00	17
grapes	1.00	1.00	1.00	14
jute	0.92	1.00	0.96	23
kidneybeans	1.00	1.00	1.00	20
lentil	1.00	1.00	1.00	11
maize	1.00	1.00	1.00	21
mango	1.00	1.00	1.00	19
mothbeans	1.00	1.00	1.00	24
mungbean	1.00	1.00	1.00	19
muskmelon	1.00	1.00	1.00	17
orange	1.00	1.00	1.00	14
papaya	1.00	1.00	1.00	23
pigeonpeas	1.00	1.00	1.00	23
pomegranate	1.00	1.00	1.00	23
rice	1.00	0.89	0.94	19
watermelon	1.00	1.00	1.00	19
accuracy			1.00	440
macro avg	1.00	1.00	1.00	440
weighted avg	1.00	1.00	1.00	440

4. K- Nearest Neighbour

- KNN can be used for both classification and regression predictive problems. However, it is more widely used in classification problems in the industry.

- The number of neighbors is the core deciding factor
  .
- KNN has the following basic steps:
  1.Calculate distance
  2.Find closest neighbors
  3.Vote for labels

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	23
banana	1.00	1.00	1.00	21
blackgram	0.95	0.95	0.95	20
chickpea	1.00	1.00	1.00	26
coconut	1.00	1.00	1.00	27
coffee	1.00	1.00	1.00	17
cotton	0.94	1.00	0.97	17
grapes	1.00	1.00	1.00	14
jute	0.74	1.00	0.85	23
kidneybeans	0.91	1.00	0.95	20
lentil	0.69	1.00	0.81	11
maize	1.00	0.95	0.98	21
mango	0.83	1.00	0.90	19
mothbeans	1.00	0.83	0.91	24
mungbean	1.00	1.00	1.00	19
muskmelon	1.00	1.00	1.00	17
orange	1.00	1.00	1.00	14
papaya	1.00	0.91	0.95	23
pigeonpeas	1.00	0.70	0.82	23
pomegranate	1.00	1.00	1.00	23
rice	0.92	0.63	0.75	19
watermelon	1.00	1.00	1.00	19
accuracy			0.95	440
macro avg	0.95	0.95	0.95	440
weighted avg	0.96	0.95	0.95	440

5. Random Forest

- Random forests is a supervised learning algorithm.A forest is comprised of trees. It is said that the more trees it has, the more robust a forest is.

- Random forests creates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting.

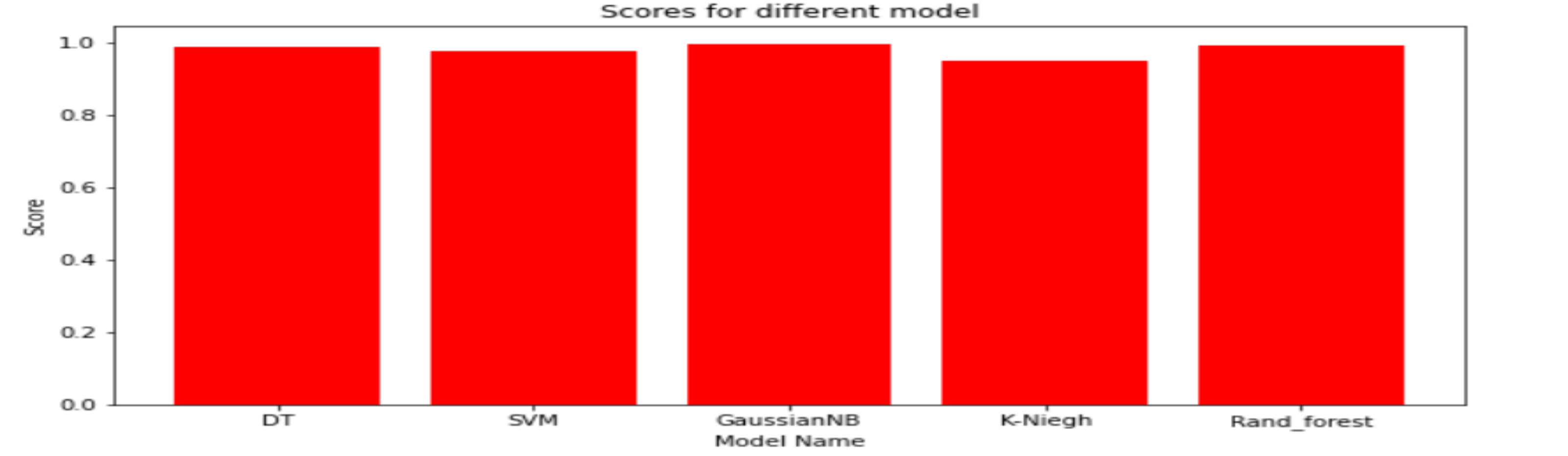
- It does not suffer from the overfitting problem. The main reason is that it takes the average of all the predictions, which cancels out the biases.

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	23
banana	1.00	1.00	1.00	21
blackgram	0.95	0.95	0.95	20
chickpea	1.00	1.00	1.00	26
coconut	1.00	1.00	1.00	27
coffee	1.00	1.00	1.00	17
cotton	0.94	1.00	0.97	17
grapes	1.00	1.00	1.00	14
jute	0.74	1.00	0.85	23
kidneybeans	0.91	1.00	0.95	20
lentil	0.69	1.00	0.81	11
maize	1.00	0.95	0.98	21
mango	0.83	1.00	0.90	19
mothbeans	1.00	0.83	0.91	24
mungbean	1.00	1.00	1.00	19
muskmelon	1.00	1.00	1.00	17
orange	1.00	1.00	1.00	14
papaya	1.00	0.91	0.95	23
pigeonpeas	1.00	0.70	0.82	23
pomegranate	1.00	1.00	1.00	23
rice	0.92	0.63	0.75	19
watermelon	1.00	1.00	1.00	19
accuracy			0.95	440
macro avg	0.95	0.95	0.95	440
weighted avg	0.96	0.95	0.95	440

CONCLUSION AND DISCUSSION

```
# creating dictionarries for scores
total_score = {"DT":score_dt, "SVM":score_svm, "GaussianNB":score_nb, "K-Niegh":score_kn, "Rand_forest":score_rf}
total_score
```

```
{'DT': 0.9863636363636363,
 'GaussianNB': 0.9954545454545455,
 'K-Niegh': 0.95,
 'Rand_forest': 0.9931818181818182,
 'SVM': 0.9772727272727273}
```



1. SOIL MANAGEMENT

- Soil is a heterogeneous natural resource, with complex processes and mechanisms that are difficult to understand.
- Soil properties allow researchers to understand the dynamics of ecosystems and the impingement in agriculture.
- Soil temperature alone plays a significant role for the accurate analysis of the climate change effects of a region and eco-environmental conditions.
- It is a significant meteorological parameter controlling the interactive processes between ground and atmosphere.
- A study presented the comparison of four regression models for the prediction of soil organic carbon (OC), moisture content (MC), and total nitrogen (TN).
- The authors used a visible-near infrared (VIS-NIR) spectrophotometer to collect soil spectra from 140 unprocessed and wet samples of the top layer of Luvisol soil types.
- The samples were collected from an arable field in Premslin, Germany in August 2013, after the harvest of wheat crops.
- They concluded that the accurate prediction of soil properties can optimize soil management.

REVIEW WORKS

- ***Kamir et al.*** used ML models to ***identify the yield gap hotspots in wheat production***.
- He generated very high-resolution yield maps using data from various sources between 2009 and 2015.
- The data was collected from various sources:
  - NDVI time-series data across Australia using the MOD13Q1 data set
  - Rainfall and temperature data were collected from historic climate data at Australia bureau of metrology
  - Maps for observed grain yield were collected at source using intelligent harvesting machines.
- The dataset generated were tested with 9 ML algorithms: RF, XGBoost, Cubist, MLP, SVR, Gaussian Process, k-NN, and Multivariate Adaptive Regression Splines.
- Out of these algorithms, SVR with RBFNN outperformed other algorithms and investigators were able to achieve the yield estimate with a R^2 of 0.77and an RMSE value of 0.55 t ha−1.
- The results were validated using 10-fold cross-validation techniques applied to the full data set.

- ***Aghighi et al.*** used various advanced regression algorithms to ***predict the yield of silage maize crops.***
- He selected maize fields located at Moghan Agro-Industrial and Animal Husbandry Company (MAIAHC), which is about 28,000 hectares' area and located in Iran.
- The crop yield dataset was collected for around 40 silage maize fields were collected for a period from 2013-2015. Additionally, he also gathered time-series NDVI data from Landsat 8 OLI satellite.
- The data was fed to advanced regression algorithms:
    a. Gaussian Process Regression
    b. SVR,
    c. Boosted Regression tree
    d. RF Regression models
- Boosted regression tree reported best evaluation parameters with an average R-value of higher than 0.87, and RMSE in a range of 8.5 to 11.10, with a mean value of 9.5 during the period 2013-14.

2. DISEASE AND WEED DETECTION

- Disease fungi, microorganisms, and bacteria take their energy from the plants they live on, which in turn affects the crop yield.
- Crop diseases constitute a major threat in agricultural production systems that deteriorate yield quality and quantity at production, storage, and transportation level. At farm level, reports on yield losses, due to plant diseases, are very common.
- Furthermore, crop diseases pose significant risks to food security at a global scale.
- If not detected at the right time may account for a huge economic loss to farmers. A lot of financial burden goes to a farmer in the form of pesticides, to get rid of diseases and restore the functioning of crops.
- Excessive use of pesticides also leads to environmental damage and the effects of the water and soil cycle of the agricultural land.
- Using an optimally designed AI system during crop growth period not only reduces the risk of crop disease and minimizes the economic impact, but it also results in minimizing the adverse impact of unsystematic farming on the environment.

REVIEW WORKS

- ***Mohanty et al.*** employed DL methods to ***detect crop disease*** from the image dataset of plant leaves.
- He used a public database consisting of smartphone generated 54,306 images of diseased and healthy plants leaves.
- These images were resized to 256×256 pixels and were assigned 38 different class labels of crop-disease pair, and transformed into 3 datasets color, grayscale and segmented.
- The dataset was then fed to two of the most common deep CNNs: AlexNet and GoogLeNet.
- Achieved an accuracy of 99.34% for GoogLeNet, and an accuracy of 85.53% for AlexNet network.
- The results were validated using F1 score, authors achieved a mean F score of 0.9886 for GoogLeNet, and a mean F1 score of 0.9848 for AlexNet.
- ***Sambasivan and Opiyo*** used a CNN based DL model to ***detect disease in cassava crops*** for imbalanced datasets.
- They took a database of 10,000 labeled images that were pre-processed to improve the image contrast using contrast limited adaptive histogram equalization algorithm.
- The model evaluation was done using the performance metrics: confusion matrix, accuracy measure, precision measure, sensitivity, and F1 score.
- The best-case accuracy of 99.30% and the lowest accuracy was reported as 76.9%.

3. CROP QUALITY

- Crop quality is very consequential for the market and, in general, is related to soil and climate conditions, cultivation practices and crop characteristics, to name a few.
- High quality agricultural products are typically sold at better prices, hence, offering larger earnings to farmers. For instance, as regards fruit quality, flesh firmness, soluble solids content, and skin color are among the most ordinary maturity indices utilized for harvesting .
- The timing of harvesting greatly affects the quality characteristics of the harvested products in both high value crops (tree crops, grapes, vegetables, herbs, etc.) and arable crops. Therefore, developing decision support systems can aid farmers in taking appropriate management decisions for increased quality of production.
- For example, selective harvesting is a management practice that may considerably increase quality
- Furthermore, crop quality is closely linked with food waste, an additional challenge that modern agriculture has to cope with, since if the crop deviates from the desired shape, color, or size, it may be thrown away
- ML algorithms combined with imaging technologies can provide encouraging results.

How ML is applied Here?

- In the study , the authors of Research paper presented and developed a new method for the detection and classification of botanical and non-botanical foreign matter embedded inside cotton lint during harvesting.
- The aim of the study was quality improvement while the minimising fiber damage.
- Another study regards pears production and, more specifically, a method was presented for the identification and differentiation of Korla fragrant pears into deciduous-calyx or persistent-calyx categories.
- The approach applied ML methods with hyperspectral reflectance imaging.
- The final study for this sub-category was by the authors of paper, in which a method was presented for the prediction and classification of the geographical origin for rice samples
- The method was based on ML techniques applied on chemical components of samples.
- More specifically, the main goal was the classification of the geographical origin of rice, for two different climate regions in Brazil; Goias and Rio Grande do Sul. The results showed that Cd, Rb, Mg, and K are the four most relevant chemical components for the classification of samples.

Results

Table 8. Crop: crop quality table.

Author	Crop	Observed Features	Functionality	Models/Algorithms	Results
[94]	Cotton	Short wave infrared hyperspectral transmittance images depicting cotton along with botanical and non-botanical types of foreign matter	Detection and classification of common types of botanical and non-botanical foreign matter that are embedded inside the cotton lint	SVM	According to the optimal selected wavelengths, the classification accuracies are over 95% for the spectra and the images.
[95]	Pears	Hyperspectral reflectance imaging	Identification and differentiation of Korla fragrant pears into deciduous-calyx or persistent-calyx categories	SVM/SPA-SVM	Deciduous-calyx pears: 93.3% accuracy Persistent-calyx pears: 96.7% accuracy
[96]	Rice	Twenty (20) chemical components that were found in composition of rice samples with inductively coupled plasma mass spectrometry	Prediction and classification of geographical origin of a rice sample	EL/RF	93.83% accuracy

REFERENCES

1. Kevin Tom Thomas, Varsha S, Merin Mary Saji, Lisha Varghese, Er. Jinu Thomas. Crop Prediction Using Machine Learning(2020). International Journal of Future Generation Communication and Networking Vol. 13, No. 3, (2020), pp. 1896–1901 https://sersc.org/journals/index.php/IJFGCN/article/download/28189/15590/
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. A. Sharma, A. Jain, P. Gupta and V. Chowdary, "Machine Learning Applications for Precision Agriculture: A Comprehensive Review," in IEEE Access, vol. 9, pp. 4843-4873, 2021, doi: 10.1109/ACCESS.2020.3048415. https://ieeexplore.ieee.org/document/9311735
4. Predicting Crops Yield: Machine Learning Nanodegree Capstone Project https://towardsdatascience.com/predicting-crops-yield-machine-learning-nanodegree-capstone-project-e6ec9349f69

THANK YOU