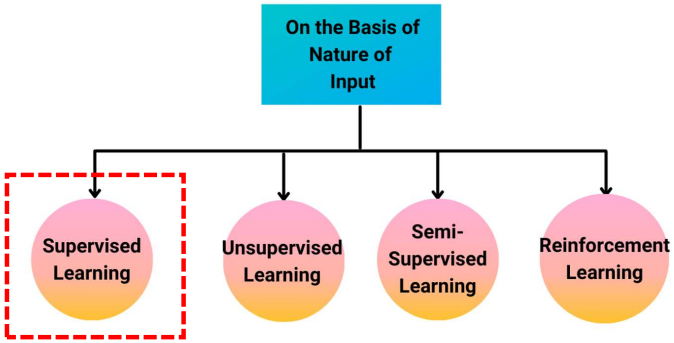
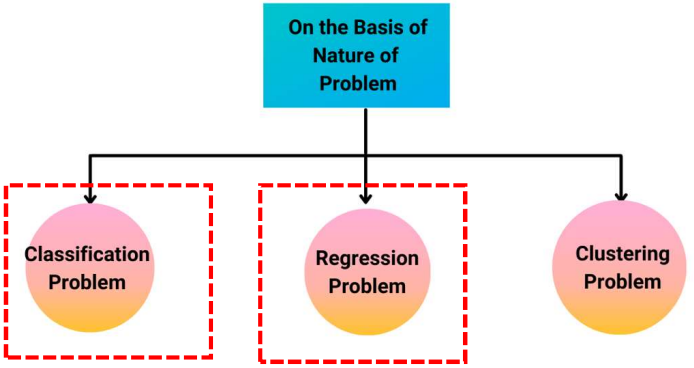
AST 426 : Data Analytics and Machine Learning in Agriculture II

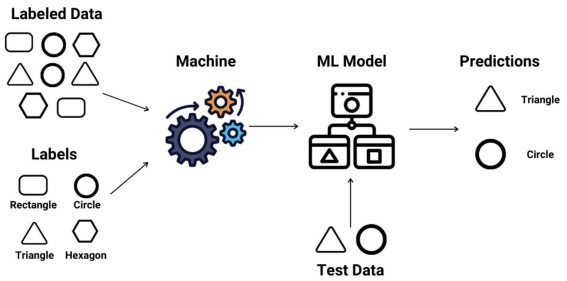
Instructor: **Pappu Kumar Yadav, Ph.D.**
Department of Agricultural & Biosystems Engineering
Machine Vision & Optical Sensors Laboratory
South Dakota State University
Fall 2024

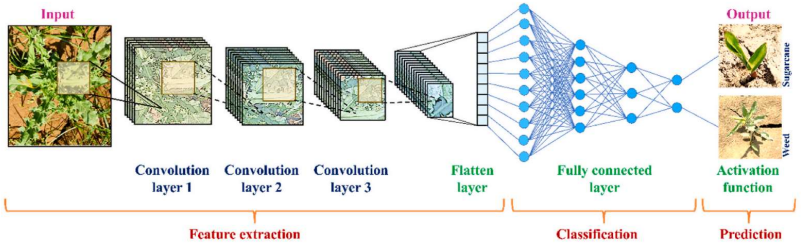
Machine Learning as a Tool for Data Analytics [Recap]

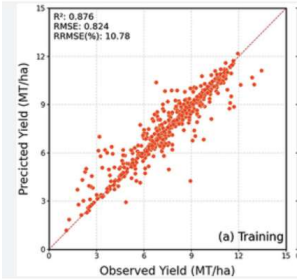
Machine Learning Types









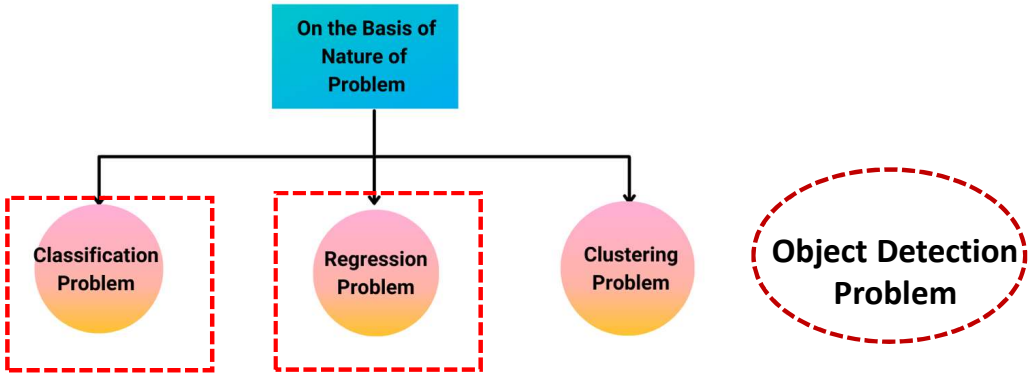


https://www.enjoyalgorithms.com/blog/classification-of-machine-learning-models

Modi, R. U., Kancheti, M., Subeesh, A., Raj, C., Singh, A. K., Chandel, N. S., ... & Singh, S. (2023). An automated weed identification framework for sugarcane crop: a deep learning approach. Crop Protection, 173, 106360.

Machine Learning as a Tool for Data Analytics [Recap]

Machine Learning Types



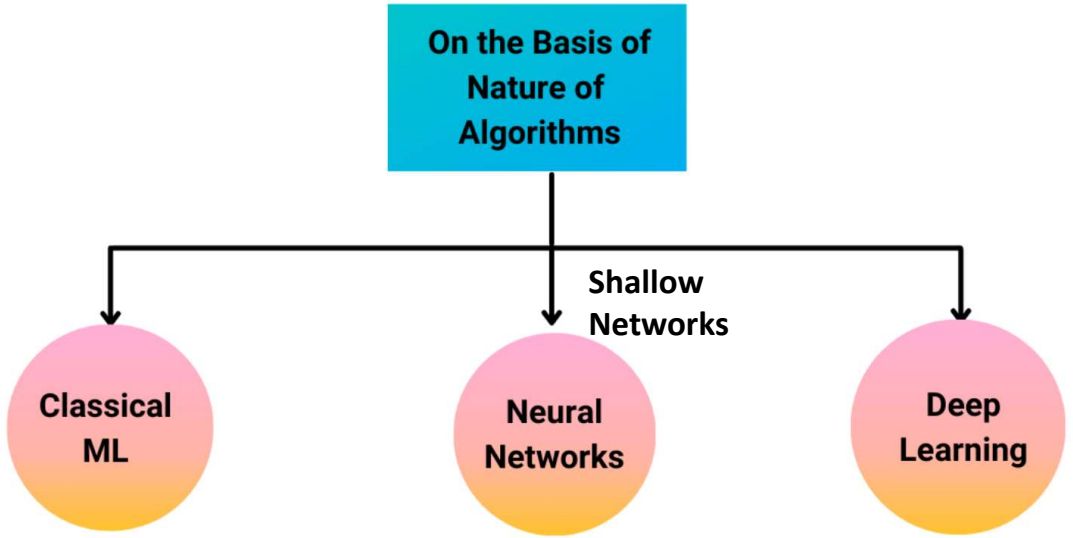
https://www.linkedin.com/posts/yolovx_ai-artificialintelligence-computervision-activity-7153748123180224513-GpaB/

https://www.enjoyalgorithms.com/blog/classification-of-machine-learning-models

Modi, R. U., Kancheti, M., Subeesh, A., Raj, C., Singh, A. K., Chandel, N. S., ... & Singh, S. (2023). An automated weed identification framework for sugarcane crop: a deep learning approach. Crop Protection, 173, 106360.

Machine Learning as a Tool for Data Analytics [Recap]

Machine Learning Types



- **Linear Regression (LR)**
- **Random Forests (RF)**
- **Support Vector Machine (SVM)**
- Decision Trees
- K-Nearest Neighbors (KNN)

- Feedforward Neural Network (FNN)
- Multi-Layer Perceptron (MLP)
- Recurrent Neural Networks (RNN)

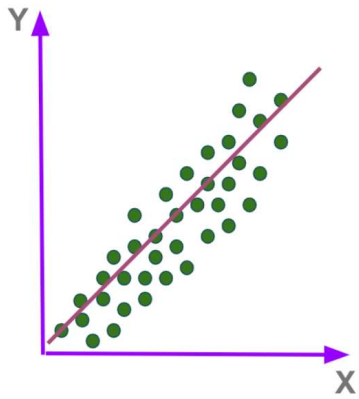
- **Convolutional Neural Networks (CNN)**
- Long Short-Term Memory (LSTM) Networks
- Generative Adversarial Networks (GANs)
- Autoencoders

https://www.enjoyalgorithms.com/blog/classification-of-machine-learning-models

Machine Learning as a Tool for Data Analytics [Recap]

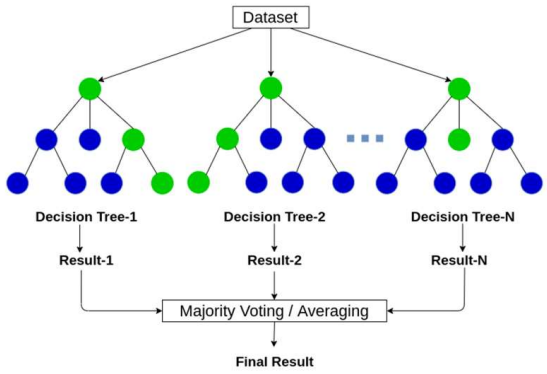
Linear Regression

- Linear regression is used to predict crop yield based on input features such as rainfall, temperature, fertilizer use, NDVI, etc. The model learns the relationship between these variables and yield, making predictions for future harvests.



Random Forests (RF)

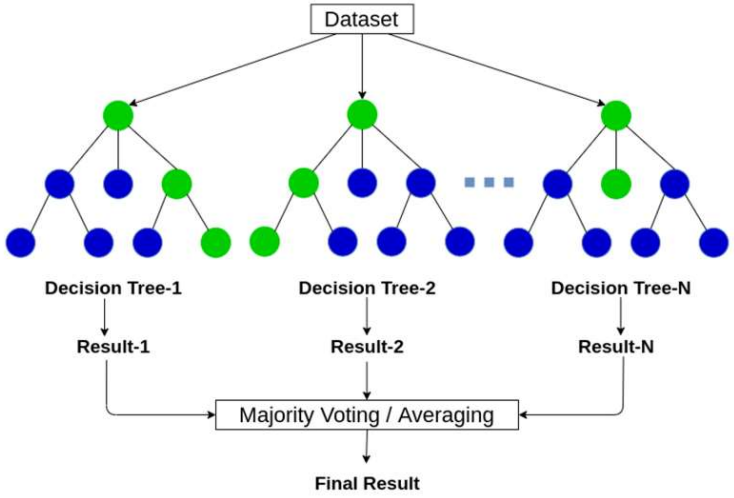
- Random forest algorithm can be trained to classify whether an image belongs to weeds by analyzing image data. It creates multiple decision trees and combines their outputs for classification tasks.
- RF is an ensemble method that **uses multiple decision trees** to make a final prediction.
- Each tree looks at different factors, and the final result is a majority vote.



Decision and Classification Trees, Clearly Explained!!! (youtube.com)

Machine Learning as a Tool for Data Analytics

Random Forests (RF)



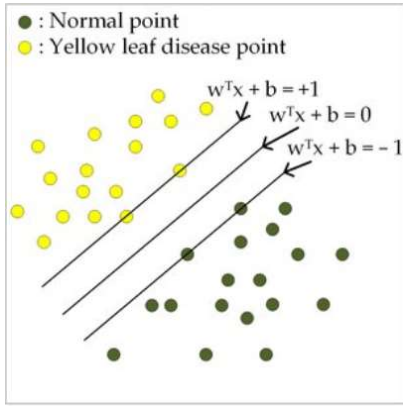
What is Random Forest? (youtube.com)

Random Forest Algorithm | Random Forest Complete Explanation | Data Science Training | Edureka (youtube.com)

Machine Learning as a Tool for Data Analytics

Support Vector Machine (SVM)

- **SVM can classify images of leaves as healthy or diseased** by identifying the optimal boundary between the two classes in the feature space.
- SVM finds the best boundary (hyperplane) to separate different categories.
- It can classify types of crops (corn, soybean) based on satellite/drone images by drawing a boundary between the two classes.
- It's **often used for binary classification** problems like distinguishing between diseased and healthy crops.



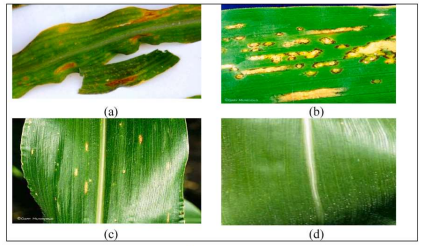
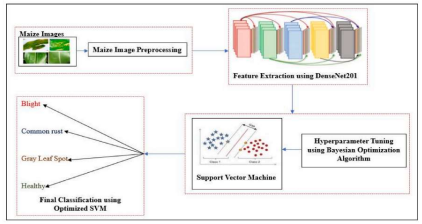


Fig. 1. Samples of Maize Leaf (a)Blight (b) common rust (c) Gray Leaf Spot (d) Healthy.



Dash, A., Sethy, P. K., & Behera, S. K. (2023). Maize disease identification based on optimized support vector machine using deep feature of DenseNet201. Journal of Agriculture and Food Research, 14, 100824.

Machine Learning as a Tool for Data Analytics

Support Vector Machine (SVM)

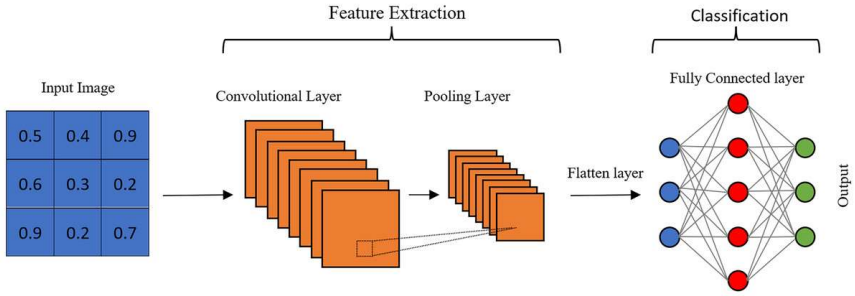
Support Vector Machine (SVM) in 2 minutes (youtube.com)

Support Vector Machines: All you need to know! (youtube.com)

Machine Learning as a Tool for Data Analytics

Convolutional Neural Network (CNN)

- Convolutional Neural Networks (CNNs) are a class of deep learning models commonly used for image classification, object detection, and pattern recognition.
- Apply filters to the input images to create feature maps, which detect features like edges, textures, and patterns.
- CNNs are used to analyze images of plants to detect diseases like blight or mildew. The model learns the distinguishing features of healthy and diseased plants, such as leaf color, texture, and shape, through its convolutional layers.
- Some common CNN architectures used in agricultural applications: **AlexNet, YOLO (You Only Look Once)(YOLOv3, YOLOv5, YOLOv8, YOLOv10, etc.), Faster R-CNN , VGGNet (VGG16, VGG19), Mask R-CNN, ResNet,** etc.



CNN Explainer (poloclub.github.io)

Machine Learning as a Tool for Data Analytics

Convolutional Neural Network (CNN)

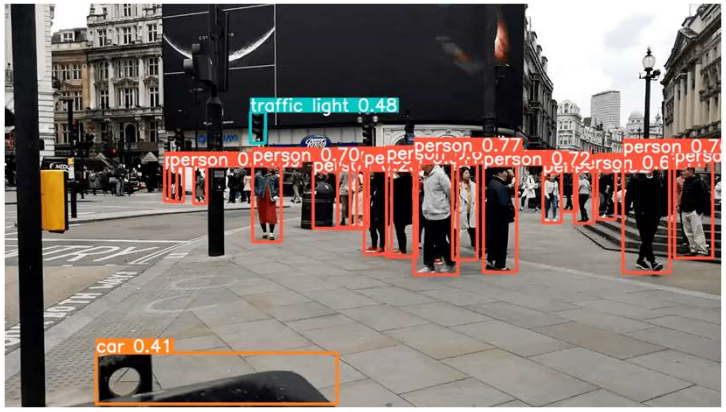
Convolutional Neural Networks (CNNs) explained (youtube.com)

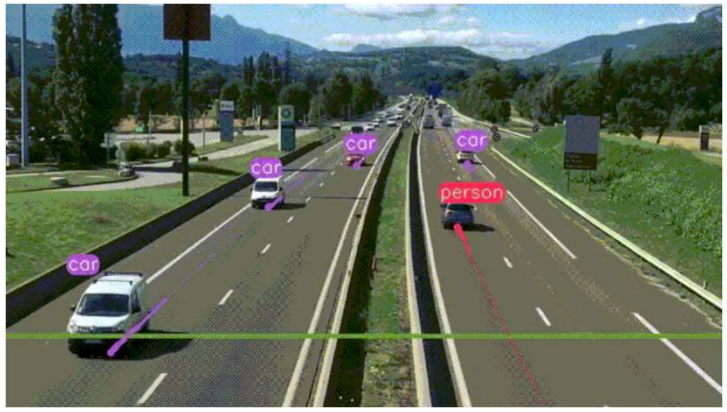
Convolutional Neural Networks Explained (CNN Visualized) (youtube.com)

Machine Learning as a Tool for Data Analytics

You Only Look Once version 8 (YOLOv8)

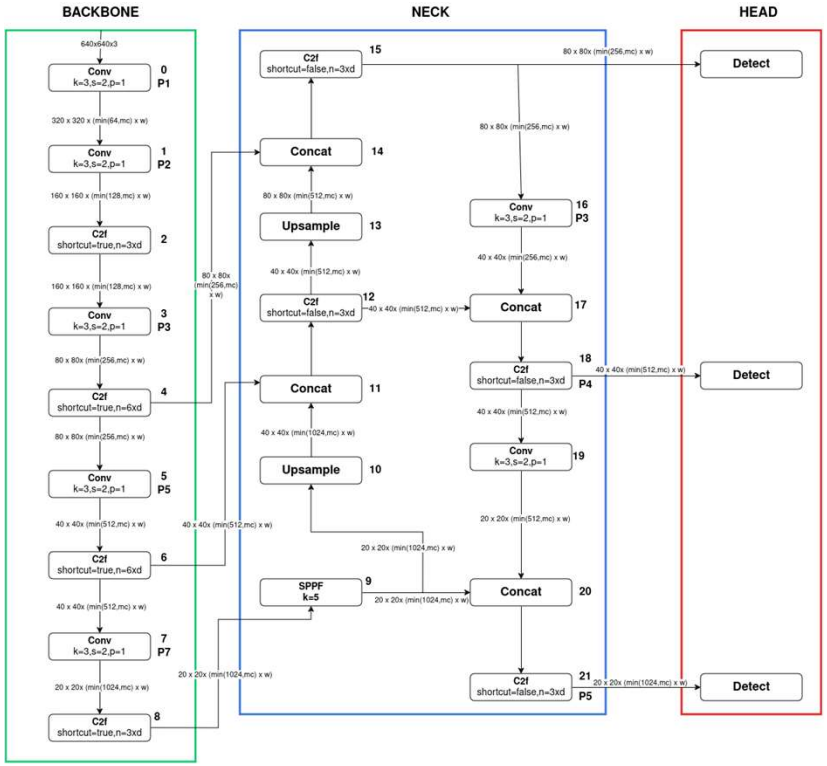
- YOLOv8 is **CNN-based deep learning model** for **object detection** that predicts bounding boxes and classifies objects in real-time.
- It **can be used to identify weeds in a crop field**, automatically locating their position and type for **precision spraying**.





Machine Learning as a Tool for Data Analytics

You Only Look Once version 8 (YOLOv8)



https://abintimilsina.medium.com/yolov8-architecture-explained-a5e90a560ce5

Machine Learning as a Tool for Data Analytics

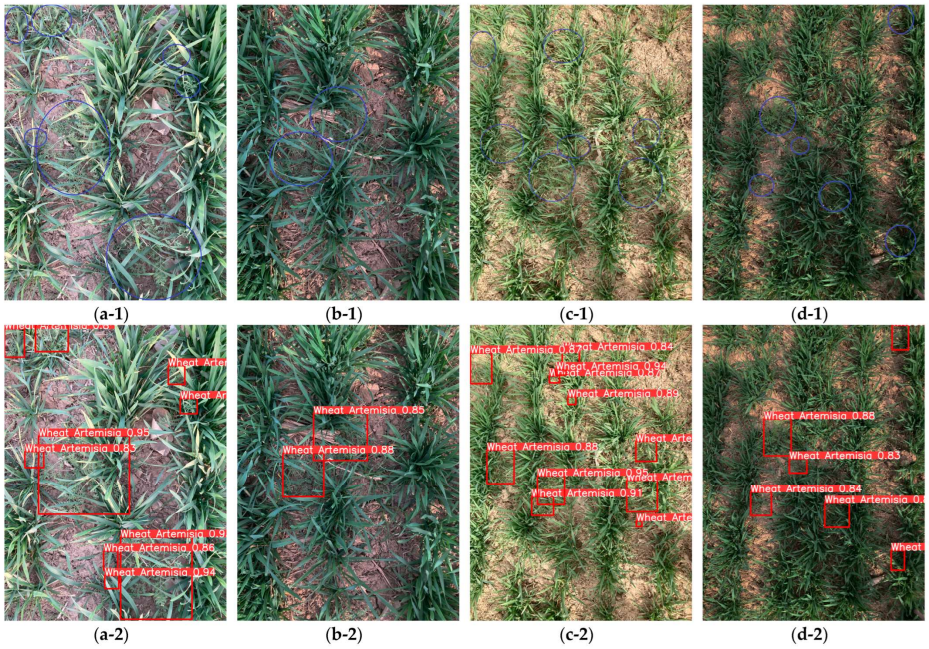
You Only Look Once version 8 (YOLOv8)

- YOLOv8 has **five variants** based on size of the network
  i. Nano : YOLOv8n
  ii. Small:YOLOv8s
  iii. Medium:YOLOv8m
  iv. Large:YOLOv8l
  v. Extra-Large:YOLOv8x

Machine Learning as a Tool for Data Analytics

Why CNN models like YOLOv8 are Powerful in Agriculture?

- CNNs are well-suited for processing high-resolution agricultural images, such as those from drones or satellites.
- No need for manual feature selection (like color or texture); CNNs automatically learn which features are important.
- Can process images quickly for real-time applications like precision spraying or automated harvesting.



Liu, Y., Zeng, F., Diao, H., Zhu, J., Ji, D., Liao, X., & Zhao, Z. (2024). YOLOv8 Model for Weed Detection in Wheat Fields Based on a Visual Converter and Multi-Scale Feature Fusion. Sensors (Basel, Switzerland), 24(13).

Machine Learning as a Tool for Data Analytics

Performance Metrics

1. Precision

- Precision measures how many of the crops predicted as diseased were diseased.
- High precision means if your model says a crop is diseased, it's very likely true, which reduces false alarms.

$$Precision = \frac{TP}{TP + FP}$$

Machine Learning as a Tool for Data Analytics

Performance Metrics

2. Recall

- Recall measures how many of the actual diseased crops were correctly identified by the model.
- High recall ensures you capture most of the diseased crops, preventing missed cases that could spread.

$$Recall = \frac{TP}{TP + FN}$$

Performance Metrics

3. F1-Score

- F1-score is the harmonic mean of precision and recall, balancing both.
- It helps find a middle ground between catching all diseased crops (recall) and ensuring accuracy in those predictions (precision).

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Machine Learning as a Tool for Data Analytics

Performance Metrics

4. Accuracy

- Accuracy measures the proportion of correct predictions out of all predictions made.
- If you are classifying healthy vs. diseased crops, accuracy tells you how many crops were correctly identified, both healthy and diseased.

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

Performance Metrics

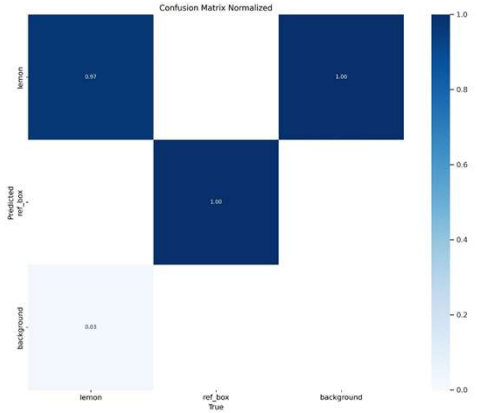
5. Mean Average Precision (mAP)

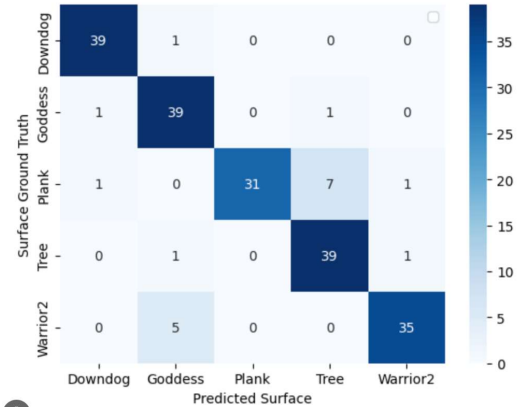
- mAP evaluates object detection models (like YOLOv8) and considers both precision and recall over different confidence thresholds.
- For detecting weeds, mAP measures how well the model identifies weeds, considering accuracy across different conditions (small weeds, large weeds).

$$mAP = \frac{1}{k} \sum_{i}^{k} AP_i$$

Confusion Matrix

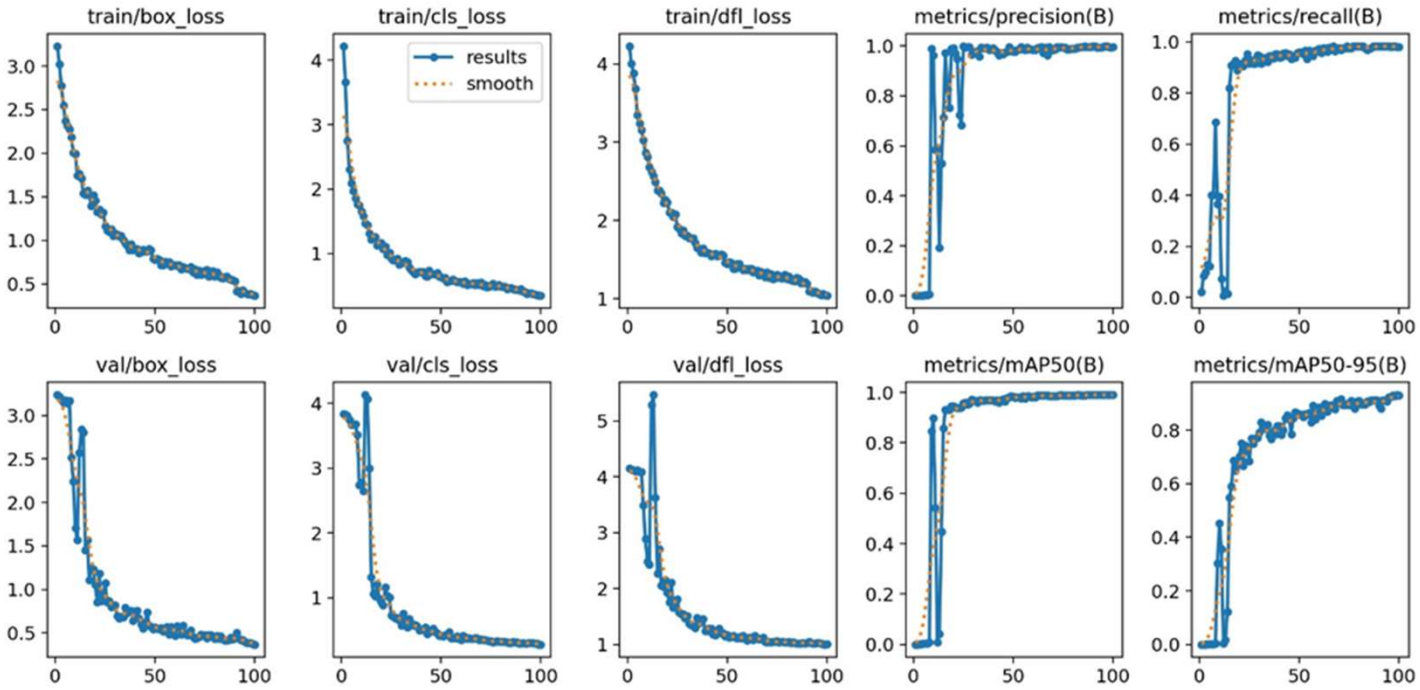
- A confusion matrix is a table used to evaluate the performance of a classification model by showing the actual vs. predicted classifications.
- It helps you understand how well your model is performing, especially for multi-class problems.





Machine Learning as a Tool for Data Analytics

YOLOv8 Training Graphs



Next Lecture

- Data Analytics and Machine Learning in Agriculture III