AST 426 :Data Analytics and Machine Learning in Agriculture I

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

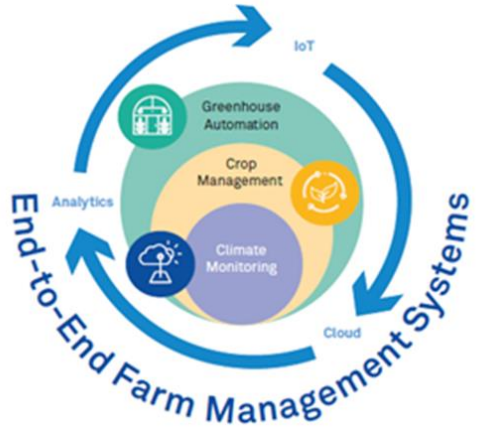
What is Data Analytics?

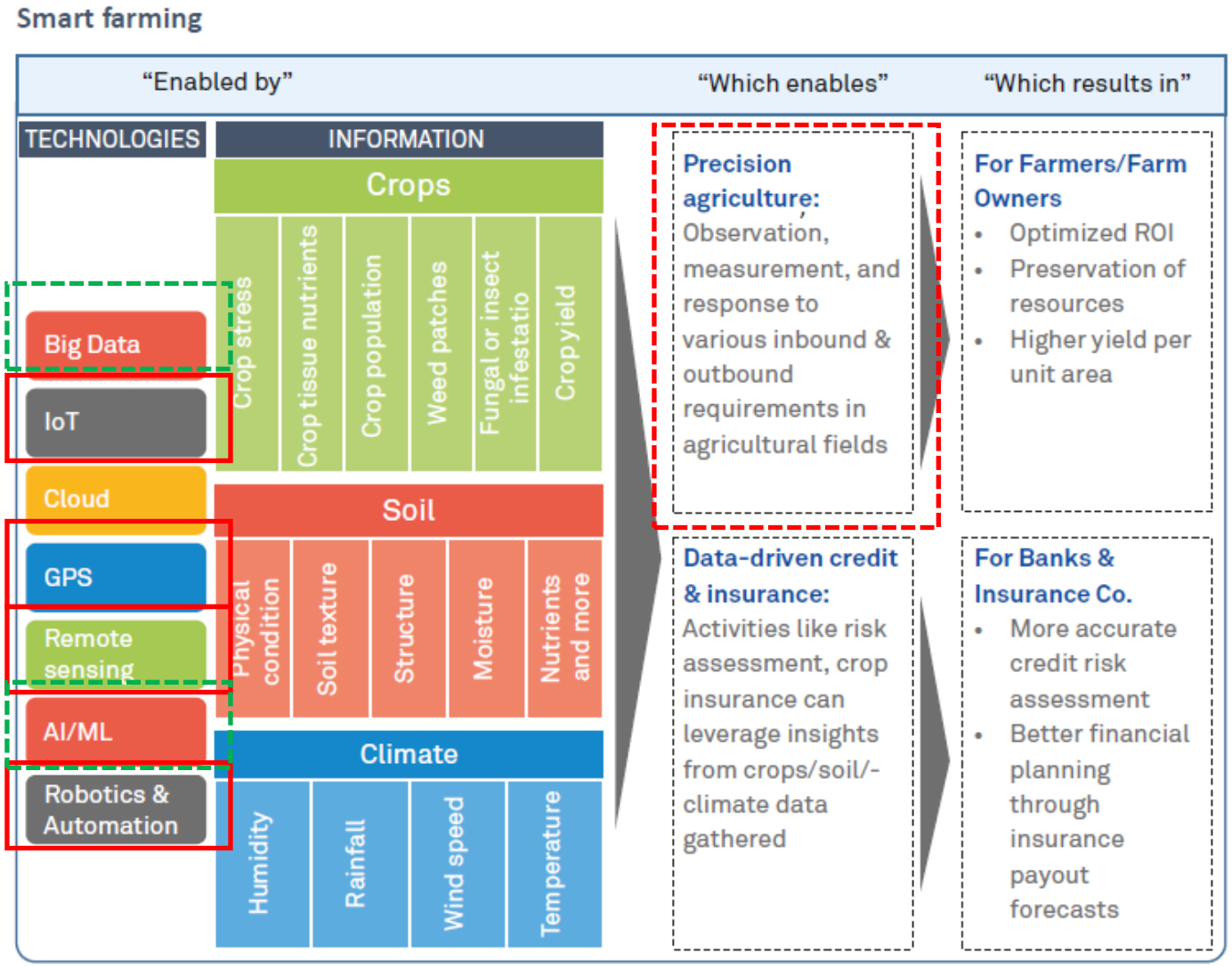
- Data analytics refers to the **science of examining raw data** with the **purpose of drawing conclusions about that information**.
- Decision making in agriculture regarding irrigation, fertigation, pesticide application, etc. are the examples of data analytics in agriculture.
- Data analytics is crucial in **precision agriculture** as it **helps optimize farming processes, reduce resource waste, and increase yield**.
- To fully realize **Smart Farming objectives**, a plethora of skills within data analytics will be necessary viz. **data capture, storage, transfer, transformation and analytics, and data-driven marketing**

https://www.wipro.com/analytics/smart-farming-powered-by-analytics/

https://iabac.org/blog/the-future-of-farming-utilizing-data-analytics-in-agriculture

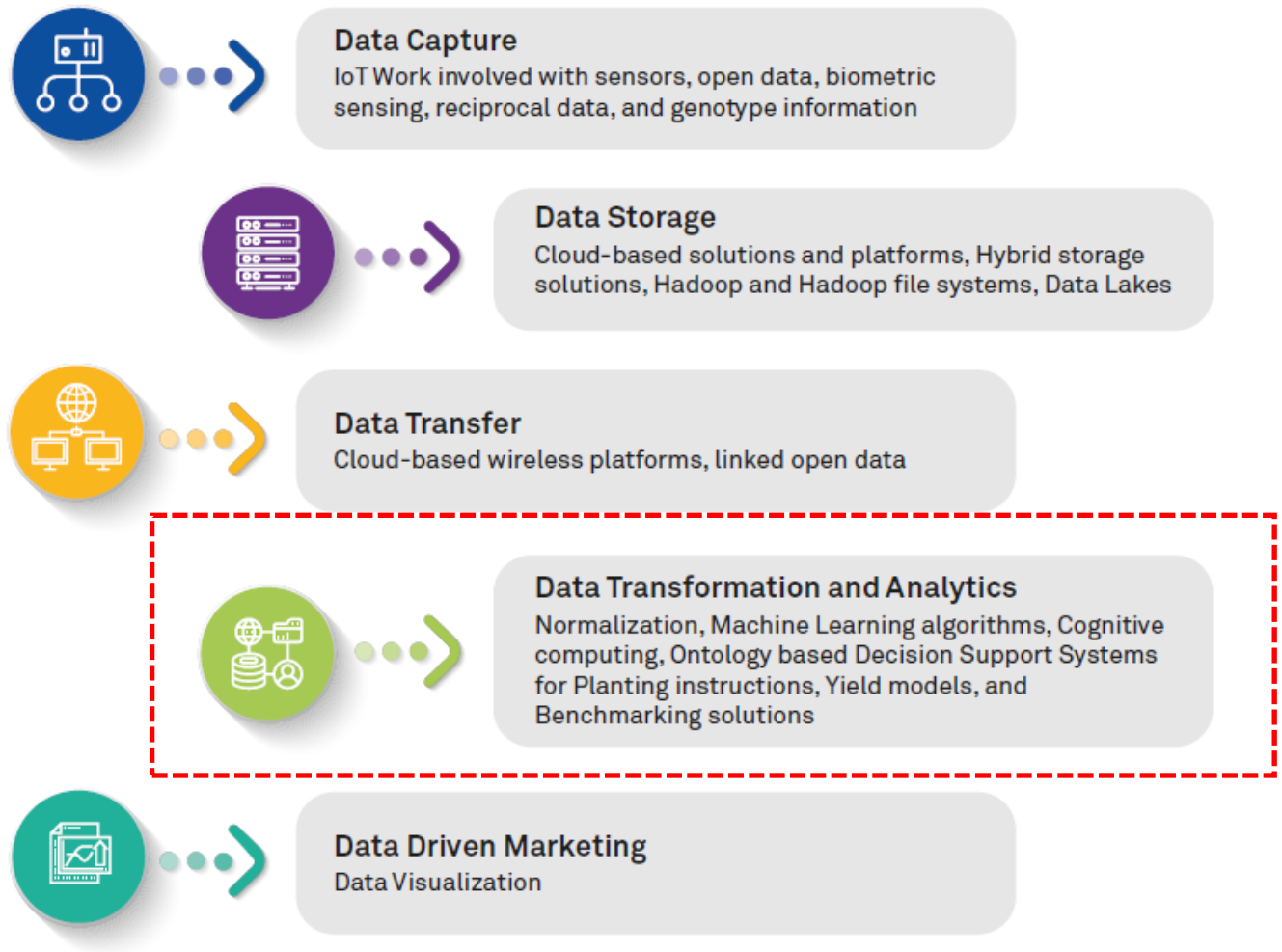
Data Analytics for Precision Agriculture and Smart Farming?





https://www.wipro.com/analytics/smart-farming-powered-by-analytics/

Data Analytics for Precision Agriculture and Smart Farming?

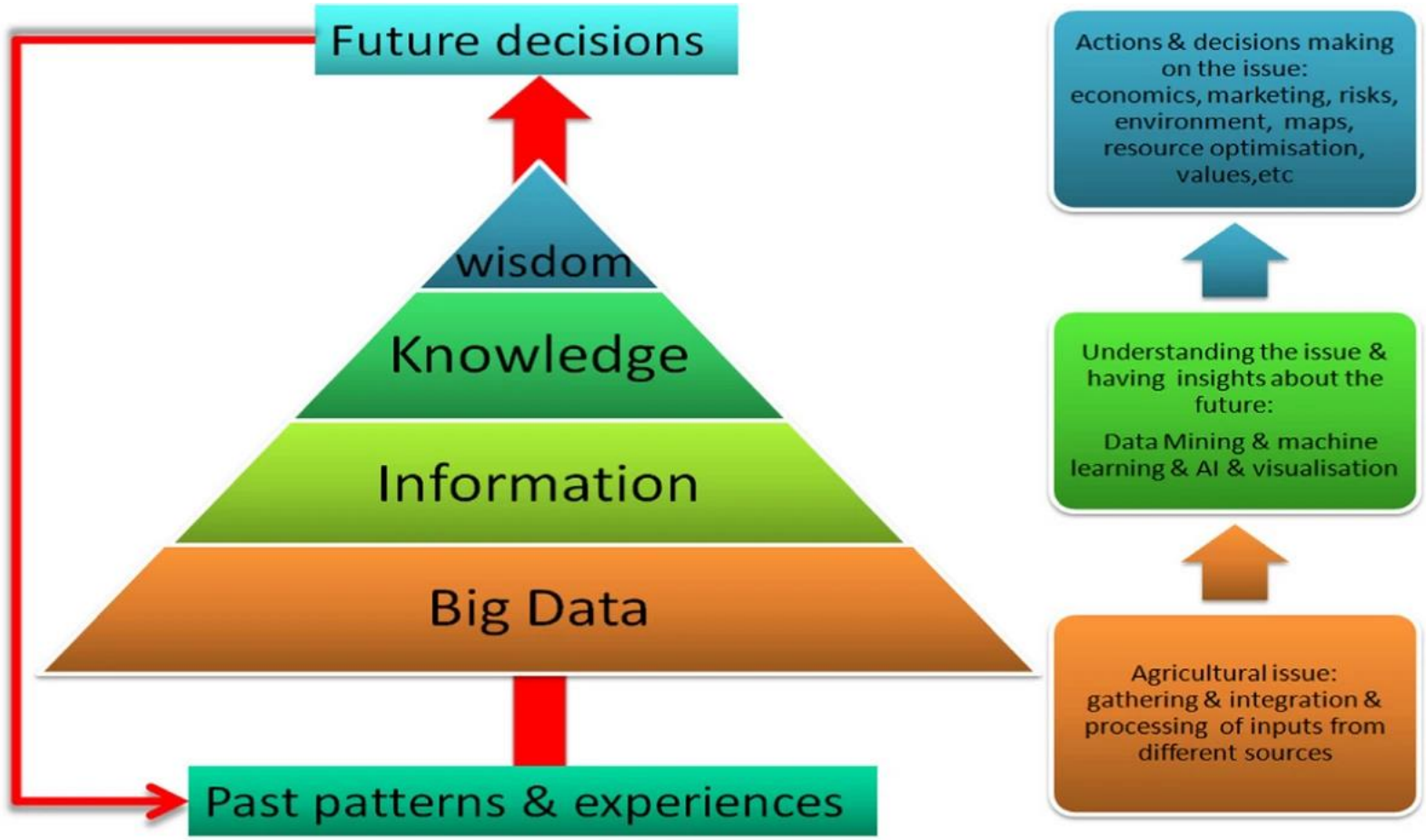


https://www.wipro.com/analytics/smart-farming-powered-by-analytics/

Role of Data Analytics in Agriculture

Chergui, N., & Kechadi, M. T. (2022). Data analytics for crop management: a big data view. Journal of Big Data, 9(1), 123.

Data Analytics Process for Crop Yield Monitoring

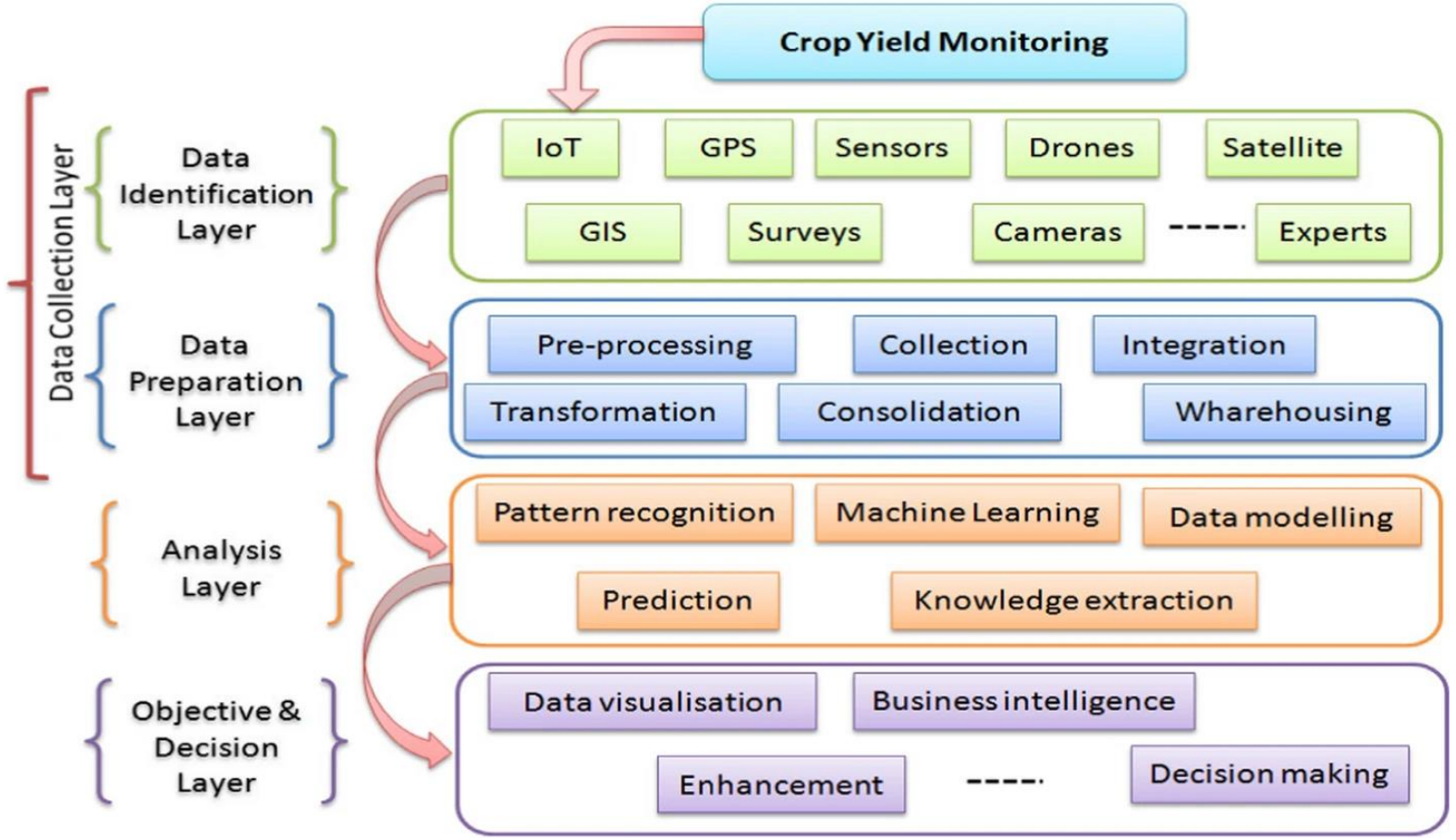


Chergui, N., & Kechadi, M. T. (2022). Data analytics for crop management: a big data view. Journal of Big Data, 9(1), 123.

Data Analytics System Architecture for Crop Yield Monitoring

Remember!

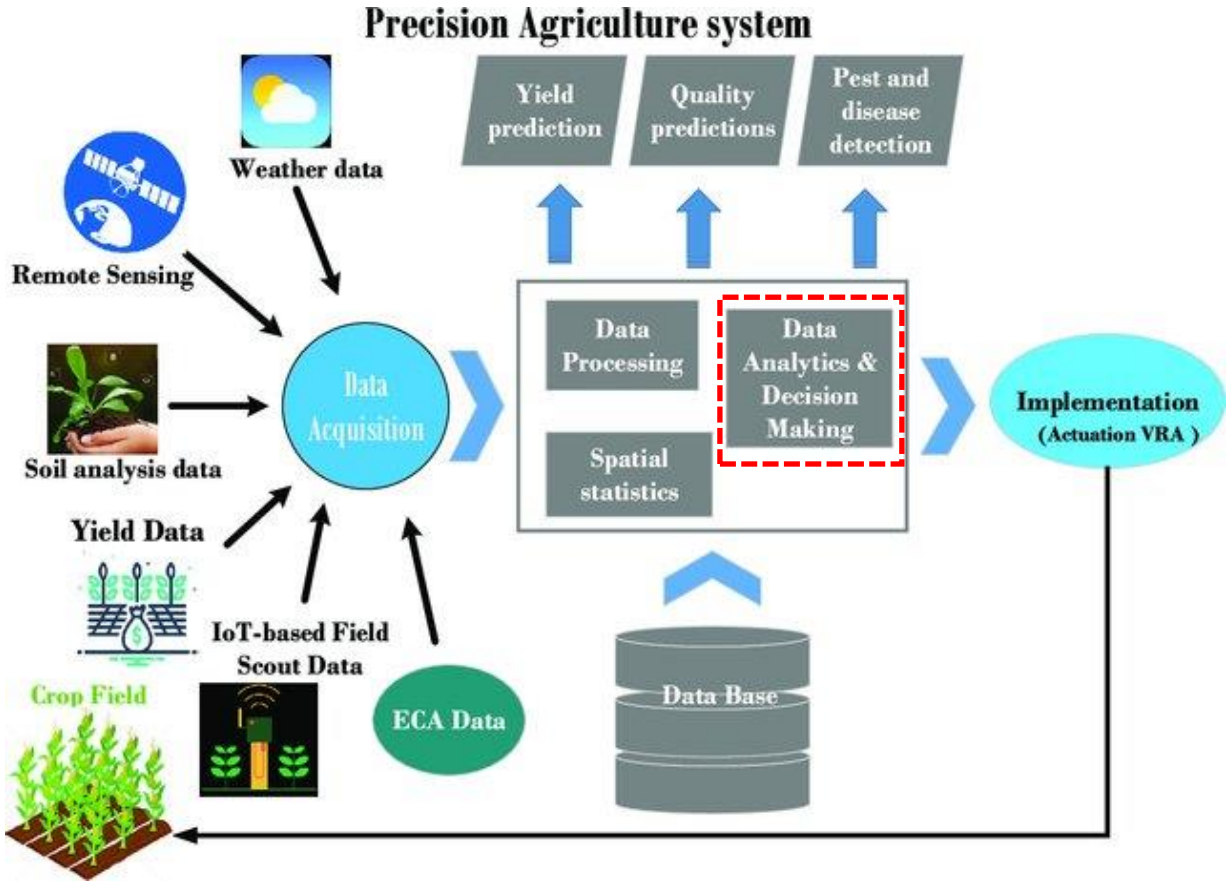
- The main goal of the DA process is the decision-making.
- Any decision should follow the state-of-the-art practice, be justifiable and scientifically sound



Chergui, N., & Kechadi, M. T. (2022). Data analytics for crop management: a big data view. Journal of Big Data, 9(1), 123.

Types of Data for Analytics in Precision Agriculture

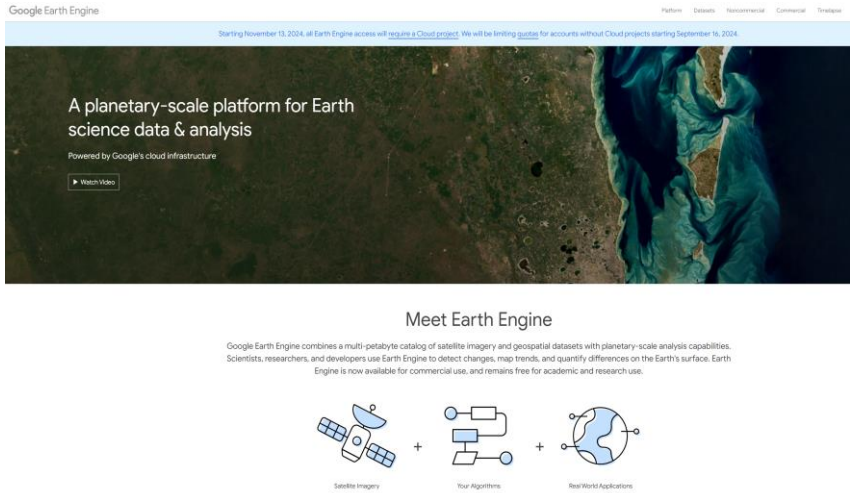
- **Weather data**: Precipitation, temperature, humidity.
- **Soil data**: Nutrients, pH, moisture levels.
- **Crop health data**: Growth stages, disease outbreaks.
- **Yield data**: Historical yield records, yield monitors.
- **Geospatial data**: Satellite and drone images.

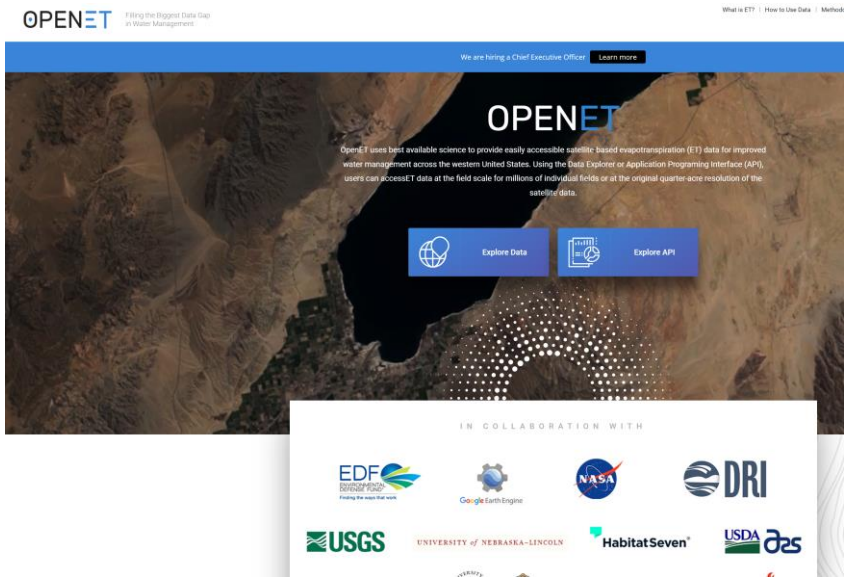


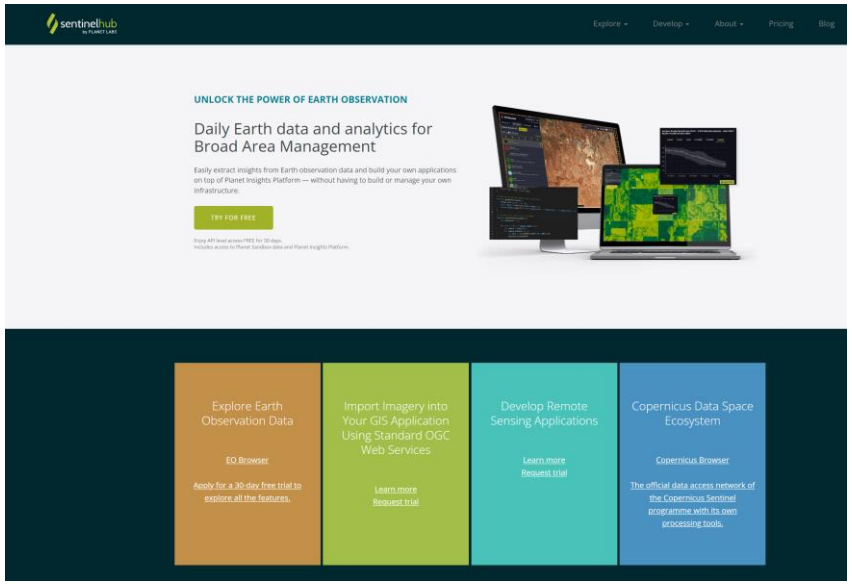
Chergui, N., & Kechadi, M. T. (2022). Data analytics for crop management: a big data view. Journal of Big Data, 9(1), 123.

Tools for Data Analytics in Precision Agriculture

- Google Earth Engine ✓
- Sentinel Hub ✓
- OpenET ✓
- Machine Learning
- ArcGIS Pro
- ETC.

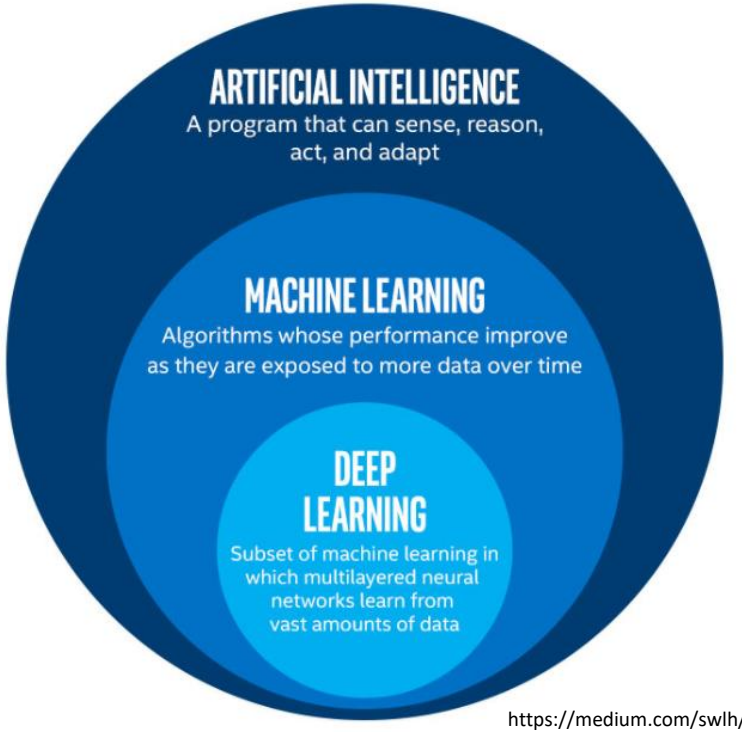






What is Machine Learning (ML)?

- Machine learning is a **subfield of artificial intelligence (AI)** that **gives computers the ability to learn without explicitly being programmed.**
- Provides machine to **imitate intelligent human behavior**
- ML is an algorithm while AI is system



https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

https://medium.com/swlh/artificial-intelligence-machine-learning-and-deep-learning-whats-the-real-difference-94fe7e528097

Machine Learning as a Tool for Data Analytics

- Machine learning **allows computer systems to continuously adjust and enhance themselves** as they accrue more “experiences.”
- The **performance of these systems can be improved by providing larger and more varied datasets** to be processed.
- While **AI is a broader field** focused on creating machines that can **mimic human intelligence**, **ML is subset of AI** that focuses on **building models that learn from data**.

Machine Learning as a Tool for Data Analytics

Do you know who used the term AI first and when? And what about ML?

- The term Artificial Intelligence (AI) was first coined by **John McCarthy in 1956**.
- **Arthur Samuel** coined the term Machine Learning (ML) in **1959**.

Machine Learning as a Tool for Data Analytics

AI component is the combination of overall components that enable the robot to sense, perceive, navigate and do tasks autonomously

The robot uses **ML component** to predict crop yields, identify crop diseases, classify weeds and crops, etc. **Uses traditional algorithms** like Random Forest, Linear Regression, Support Vector Machine (SVM)

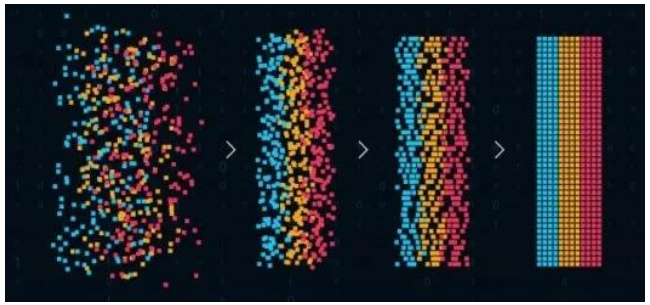
The robot uses **DL component** to identify crop diseases, weds, etc. by using **neural networks** like **convolutional neural networks (CNN)** instead of traditional ML models

Machine Learning as a Tool for Data Analytics

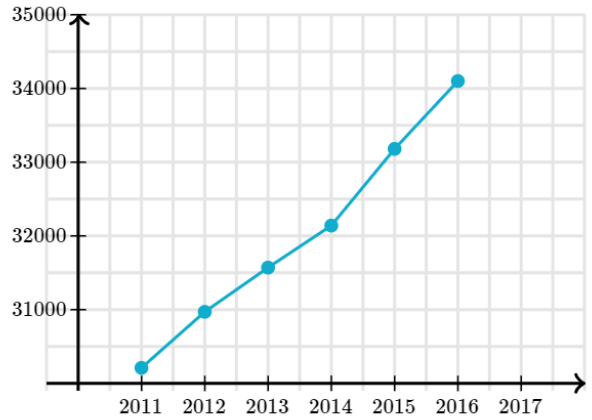
Explicit Differentiation of AI, ML, and DL Components

Aspect	AI (Artificial Intelligence)	ML (Machine Learning)	DL (Deep Learning)
Definition	The overarching system that simulates human intelligence and controls the robot's decision-making, planning, and actions.	A subset of AI that enables the robot to learn from data and make predictions or decisions based on patterns in that data.	A more specialized subset of ML that uses deep neural networks to handle complex data like images and video.
Scope	Integrates various technologies like ML, computer vision, path planning, and decision-making to manage the robot's tasks autonomously.	Focused on pattern recognition, prediction, and optimization using past data to make decisions in real-time or over time.	Focused on advanced tasks like image recognition, object detection, and real-time analysis of complex datasets.
Component Examples	- Task planning and scheduling - Navigation and obstacle avoidance - Voice command processing - Decision-making systems	- Predicting crop yield - Identifying diseases from past data - Optimizing irrigation based on soil and weather data - Classifying crops or weeds	- Detecting diseases from images (CNNs) - Real-time fruit detection - Navigating complex environments using camera data - Multi-spectral image analysis
Key Technologies	- Path planning algorithms - Rule-based systems - Natural language processing (NLP) - Decision support systems	- Random Forests - Support Vector Machines (SVM) - K-Means Clustering - Decision Trees	- Convolutional Neural Networks (CNNs) - Recurrent Neural Networks (RNNs) - Long Short-Term Memory (LSTM) networks

Role of Machine Learning in Data Analytics



- **Machine learning (ML) identifies patterns in large datasets** that might not be immediately obvious to humans. This can include identifying trends in crop growth, soil health, or market conditions based on historical data.
- **ML automates data analysis**, making it possible to **process large and complex datasets faster and more accurately than manual analysis**.
- **ML excels at making predictions based on historical data**. In agriculture, this includes yield prediction, disease outbreak forecasts, and irrigation needs.
- ML helps generate **actionable insights from data that can lead to better decision-making**. For example, recommending optimal planting schedules based on weather predictions or soil moisture data.



https://blog.devgenius.io/finding-patterns-in-large-data-sets-for-beginners-d4decfe5ca6d

Machine Learning as a Tool for Data Analytics

Types of Models in Data Analytics

1. **Mechanistic Models/Process-based Models**
   - Based on the understanding of underlying physical, biological, or chemical processes that govern a system
   - Based on established scientific principles
   - Example: Crop growth models, irrigation systems using water flow physics, plant growth simulations
2. **Empirical Models**
   - Rely on observed data to make predictions.
   - Do not necessarily explain the underlying processes but focus on fitting data to statistical relationships
   - Simpler and data-driven
   - Examples: **Linear regression to predict crop yield** based on NDVI

Machine Learning as a Tool for Data Analytics

Types of Models in Data Analytics

3. Predictive Models

- Use data to forecast future outcomes, often by learning patterns from historical data
- Make future predictions based on input data
- Widely used in ML/DL
- Example: **ML models** like Random Forest, Neural Networks, Support Vector Machine
- **Linear Regression** can also be used as a **predictive ML model**

4. Stochastic Models

- Incorporate randomness and uncertainty into the model
- Simulate a **range of possible outcomes based on probabilities**
- Examples: Weather forecasting models, pest population growth models.

Types of Models in Data Analytics

5. Deterministic Models

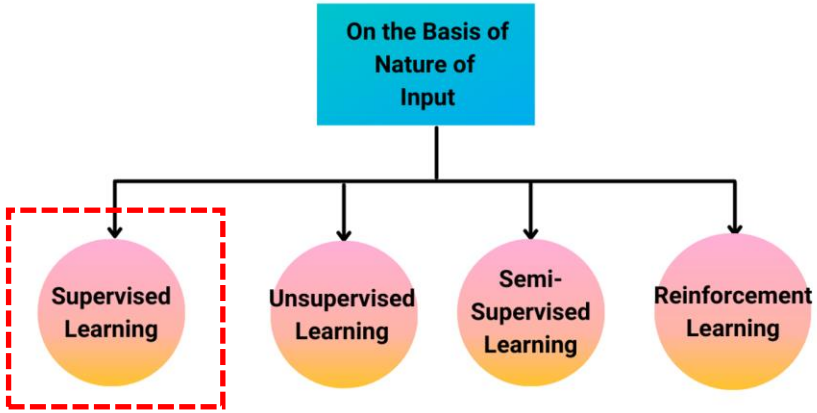
- Output is fully determined by the input parameters with no randomness involved.
- For a given set of initial conditions, the model will always produce the same result
- Example: Mathematical equations based on fixed input-output relationships (e.g., Newton's Laws)

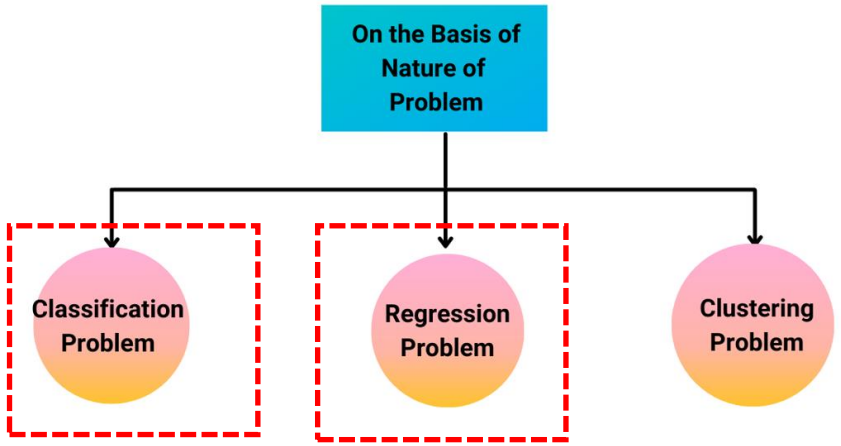
6. Hybrid Models

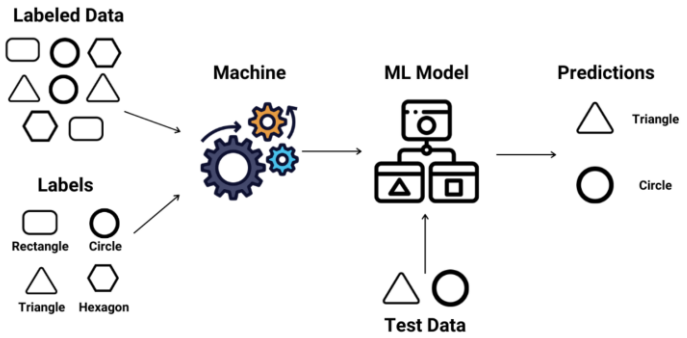
- Combine elements from multiple model types (e.g., combining mechanistic and empirical approaches).
- Use mechanistic principles for process understanding, supplemented with data-driven techniques for prediction.
- Examples: Crop growth models that use physics-based equations but refine them using machine learning predictions, **Physics Informed Neural Networks (PINNs), Scientific Machine Learning (SciML)**

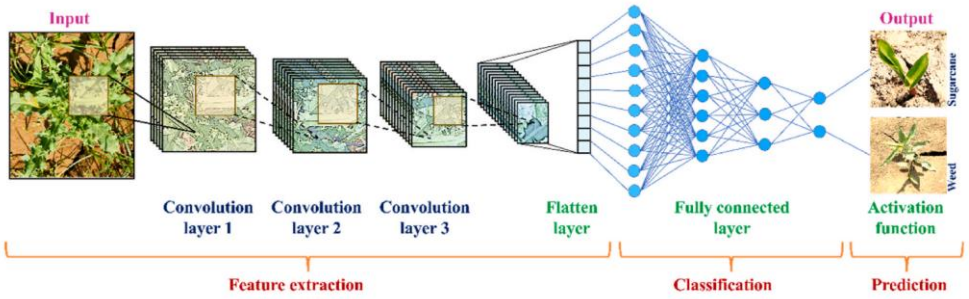
Machine Learning as a Tool for Data Analytics

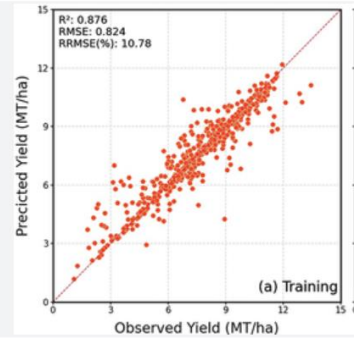
Machine Learning Types







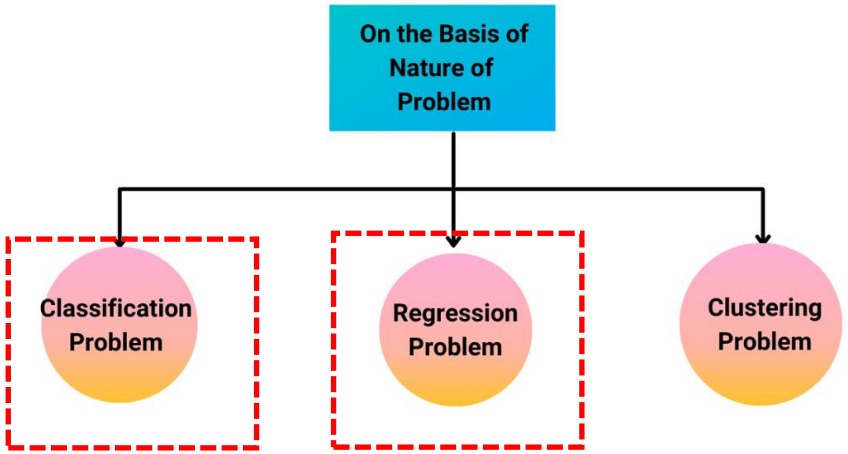




https://www.enjoyalgorithms.com/blog/classification-of-machine-learning-models

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 Types



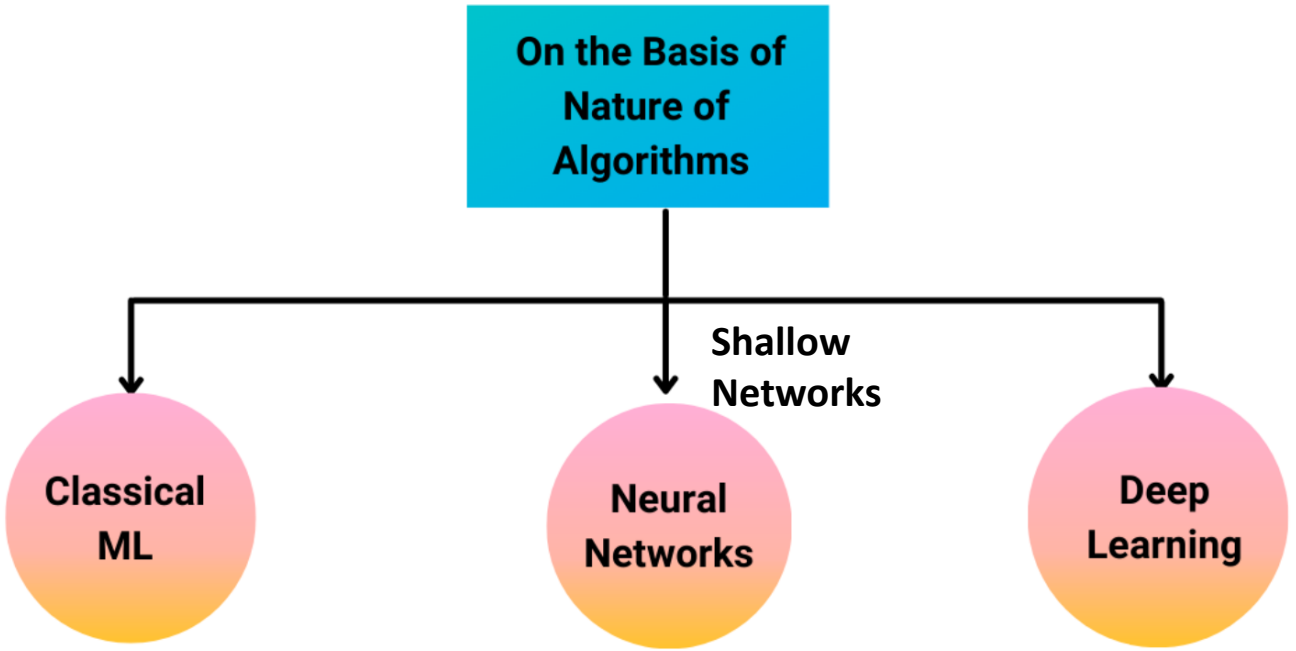
Object Detection Problem

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 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)

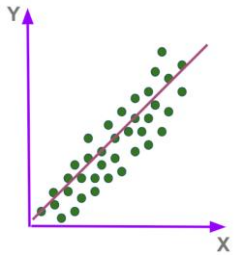
- **Convolutional Neural Networks (CNN)**
- Long Short-Term Memory (LSTM) Networks
- Generative Adversarial Networks (GANs)
- Autoencoders

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

Machine Learning as a Tool for Data Analytics

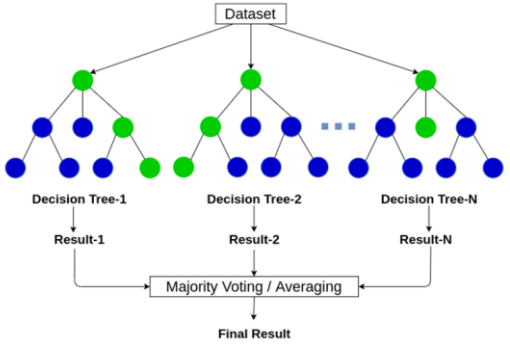
Linear Regression

- 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

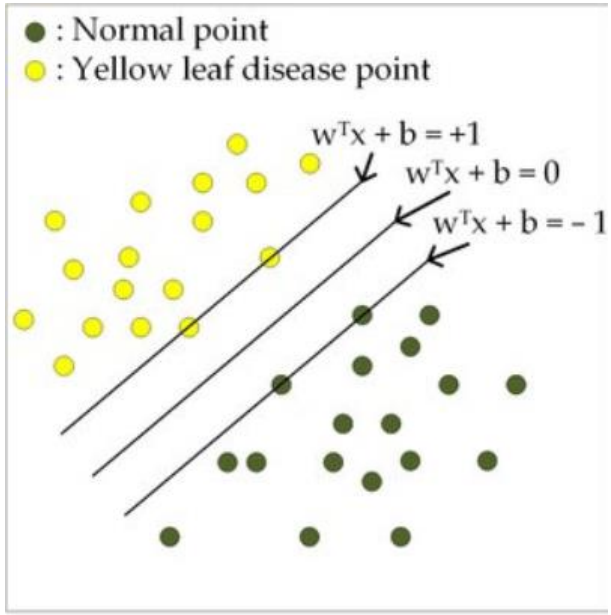
- Random forest algorithm can be trained to classify whether an image belongs to weeds by analyzing image data. It creates multiple decision trees and combines their outputs for classification tasks.



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.
- It's **often used for binary classification** problems like distinguishing between diseased and healthy crops.



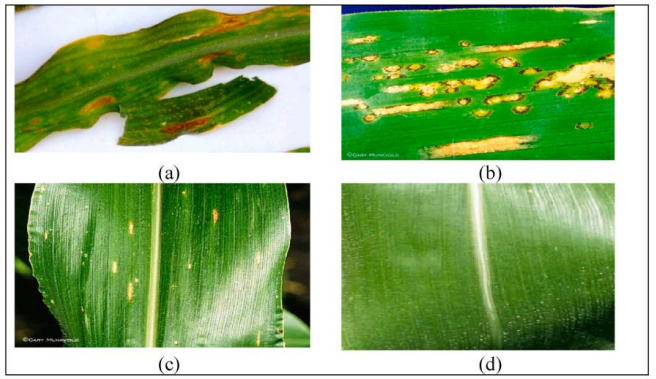
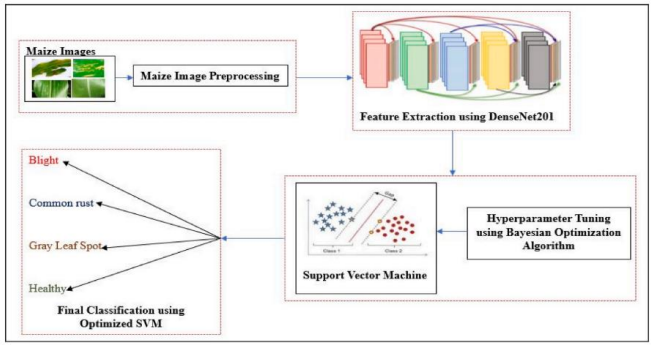


Fig. 1. Samples of Maize Leaf (a)Blight (b) common rust (c) Gray Leaf Spot (d) Healthy.

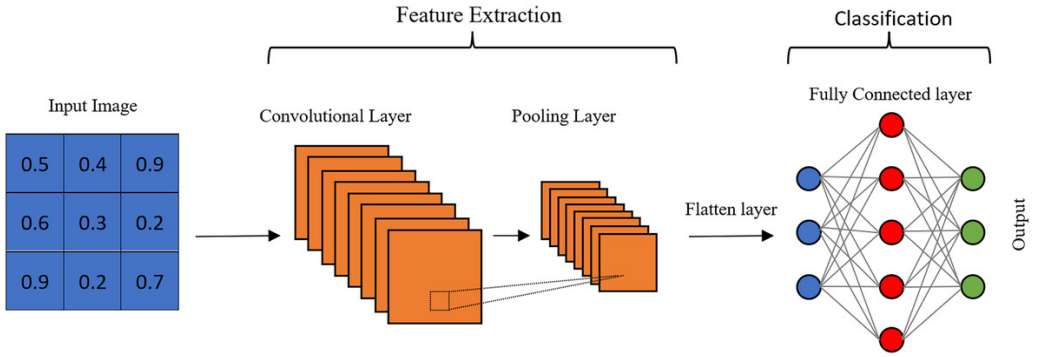


Dash, A., Sethy, P. K., & Behera, S. K. (2023). Maize disease identification based on optimized support vector machine using deep feature of DenseNet201. Journal of Agriculture and Food Research, 14, 100824.

Machine Learning as a Tool for Data Analytics

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)]

Next Lecture

- Data Analytics & Machine Learning in Agriculture II