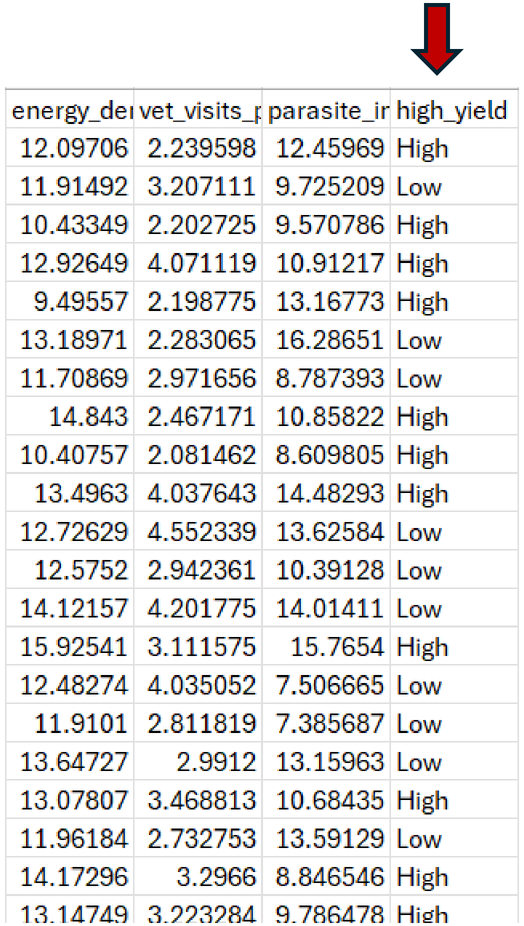
Predictive analytics

Stephen and Peace

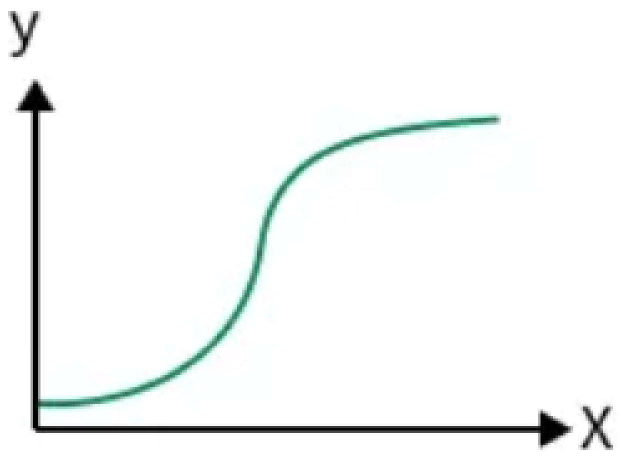
Logistic regression

- Used to solve classification problems
- Predicts the probability of a binary outcome e.g (yes/no, high/low).
- Uses a sigmoid curve for prediction
- It uses a logistic function (also called a sigmoid function) to model the probability.
- This function transforms the linear combination of input variables

into a value between 0 and 1, representing the probability of the event occurring.



energy_de	vet_visits_p	parasite_ir	high_yield
12.09706	2.239598	12.45969	High
11.91492	3.207111	9.725209	Low
10.43349	2.202725	9.570786	High
12.92649	4.071119	10.91217	High
9.49557	2.198775	13.16773	High
13.18971	2.283065	16.28651	Low
11.70869	2.971656	8.787393	Low
14.843	2.467171	10.85822	High
10.40757	2.081462	8.609805	High
13.4963	4.037643	14.48293	High
12.72629	4.552339	13.62584	Low
12.5752	2.942361	10.39128	Low
14.12157	4.201775	14.01411	Low
15.92541	3.111575	15.7654	High
12.48274	4.035052	7.506665	Low
11.9101	2.811819	7.385687	Low
13.64727	2.9912	13.15963	Low
13.07807	3.468813	10.68435	High
11.96184	2.732753	13.59129	Low
14.17296	3.2966	8.846546	High
13.14749	3.223284	9.786478	High



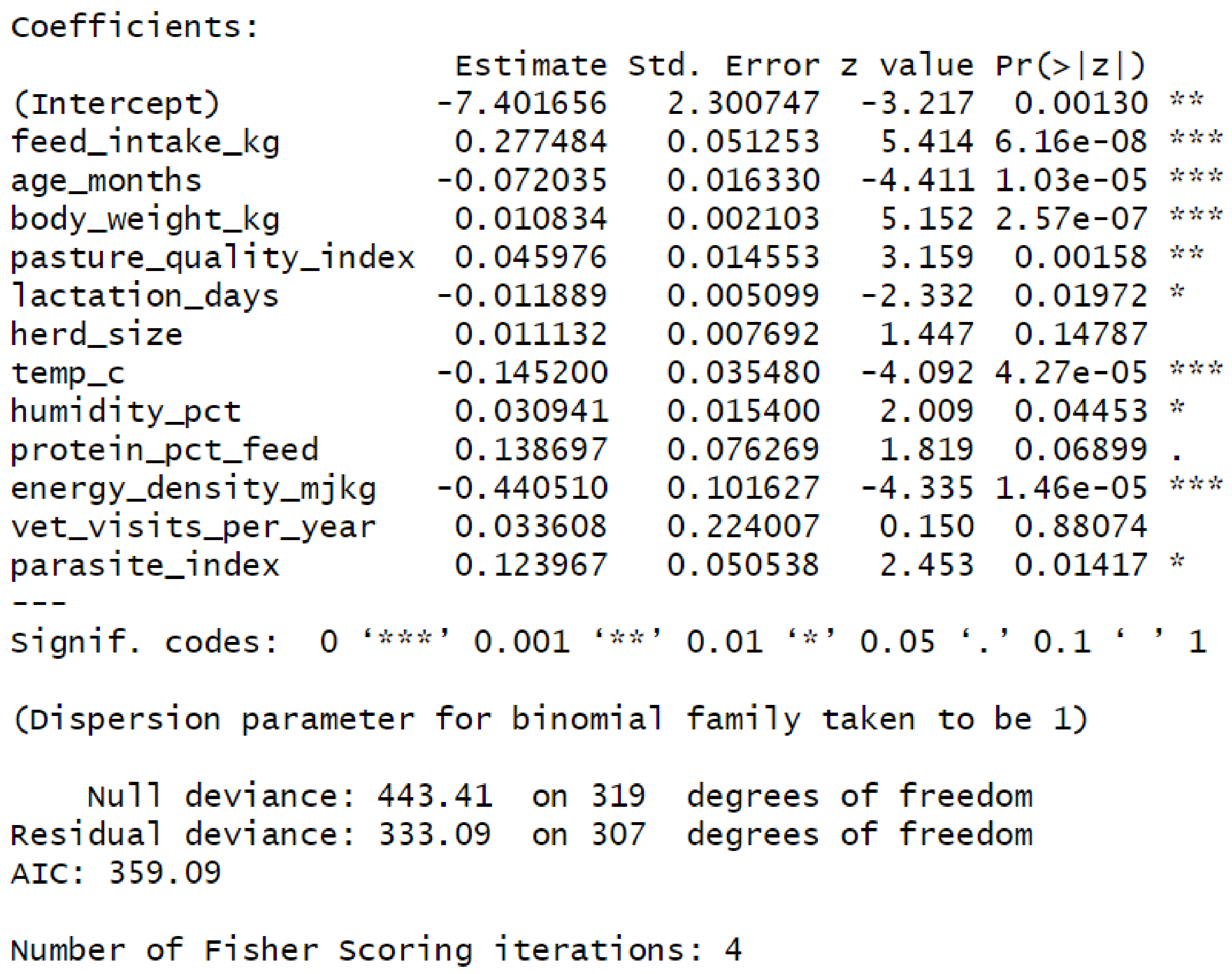
Types of Logistic Regression

- **Binomial Logistic Regression**: This type is used when the dependent variable has only two possible categories. Examples include Yes/No, Pass/Fail or 0/1. It is the most common form of logistic regression and is used for binary classification problems.

- **Multinomial Logistic Regression**: This is used when the dependent variable has three or more possible categories that are not ordered. For example, classifying animals into categories like "cat," "dog" or "sheep." It extends the binary logistic regression to handle multiple classes.

- **Ordinal Logistic Regression**: This type applies when the dependent variable has three or more categories with a natural order or ranking. Examples include ratings like "low," "medium" and "high." It takes the order of the categories into account when modeling.

Predicted yield classification

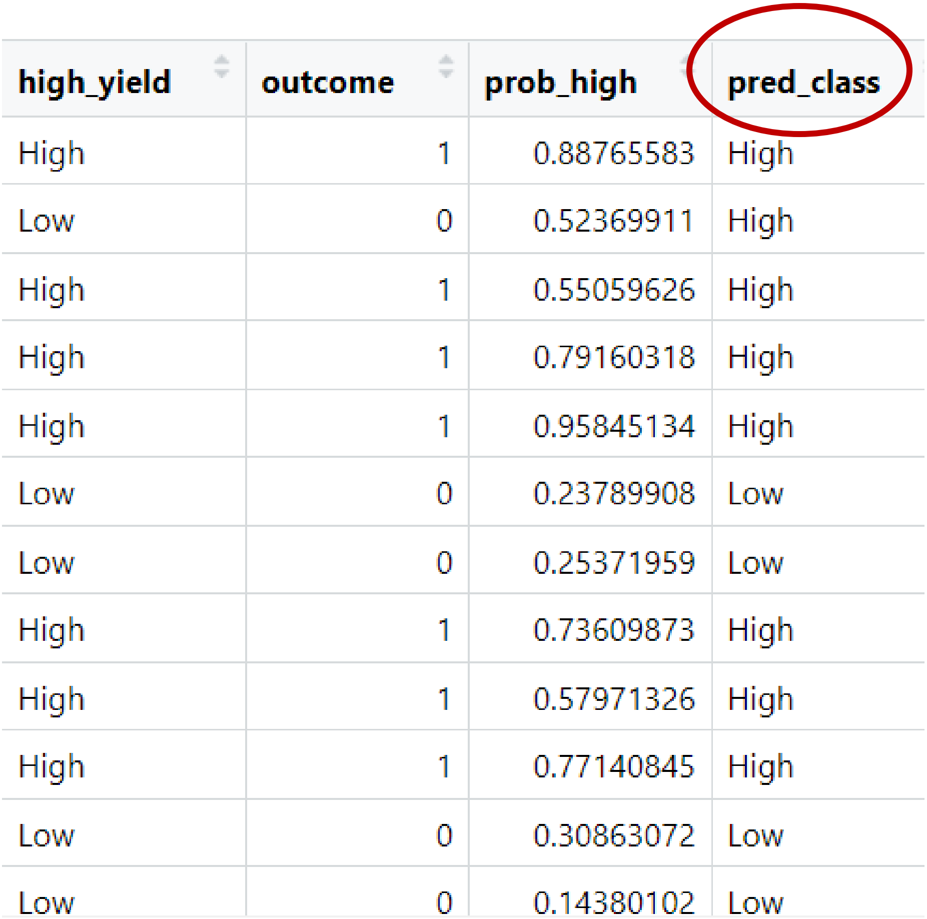


```
Coefficients:
                      Estimate Std. Error z value Pr(>|z|)    
(Intercept)          -7.401656   2.300747  -3.217  0.00130 ** 
feed_intake_kg        0.277484   0.051253   5.414 6.16e-08 ***
age_months           -0.072035   0.016330  -4.411 1.03e-05 ***
body_weight_kg        0.010834   0.002103   5.152 2.57e-07 ***
pasture_quality_index 0.045976   0.014553   3.159  0.00158 ** 
lactation_days       -0.011889   0.005099  -2.332  0.01972 *  
herd_size             0.011132   0.007692   1.447  0.14787    
temp_c               -0.145200   0.035480  -4.092 4.27e-05 ***
humidity_pct          0.030941   0.015400   2.009  0.04453 *  
protein_pct_feed      0.138697   0.076269   1.819  0.06899 .  
energy_density_mjkg  -0.440510   0.101627  -4.335 1.46e-05 ***
vet_visits_per_year   0.033608   0.224007   0.150  0.88074    
parasite_index        0.123967   0.050538   2.453  0.01417 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 443.41  on 319  degrees of freedom
Residual deviance: 333.09  on 307  degrees of freedom
AIC: 359.09

Number of Fisher Scoring iterations: 4
```



high_yield	outcome	prob_high	pred_class
High	1	0.88765583	High
Low	0	0.52369911	High
High	1	0.55059626	High
High	1	0.79160318	High
High	1	0.95845134	High
Low	0	0.23789908	Low
Low	0	0.25371959	Low
High	1	0.73609873	High
High	1	0.57971326	High
High	1	0.77140845	High
Low	0	0.30863072	Low
Low	0	0.14380102	Low

Machine Learning

Stephen and Peace

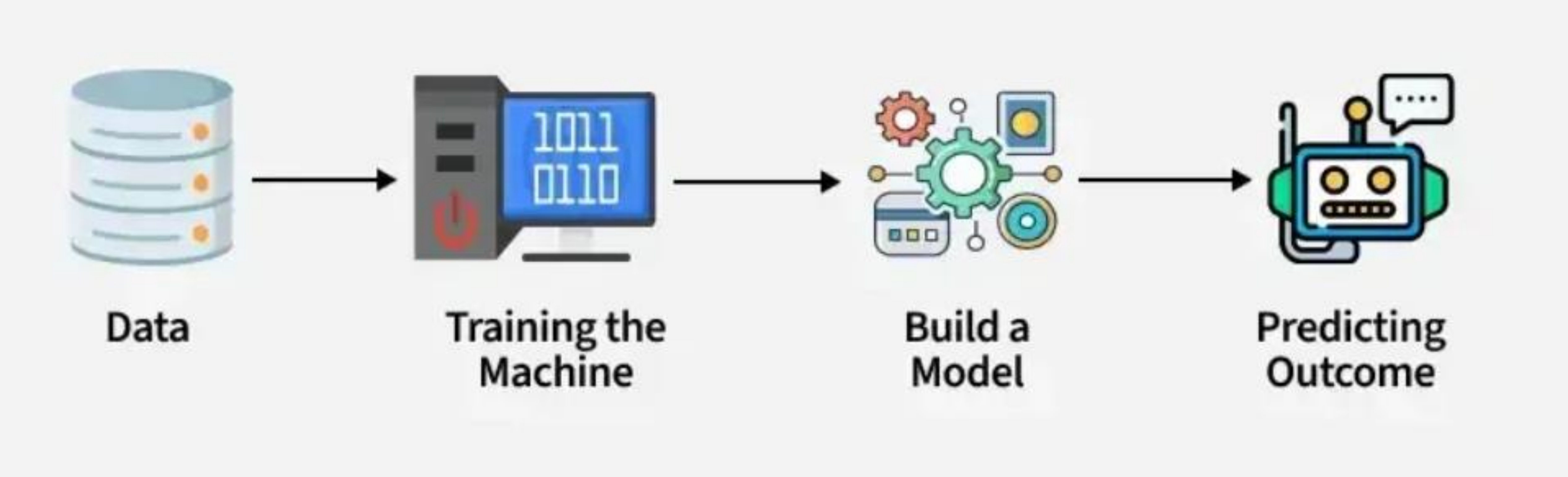
What is machine learning?

- **Machine learning (ML)** allows computers to learn and make decisions without being explicitly programmed.

- It involves feeding data into algorithms to identify patterns and make predictions on new data.

- It is used in various applications like image recognition, speech processing, language translation, recommender systems, etc.

The machine learning process



How machine learning works in R

Steps

- **Data Cleaning:** Use packages like tidyverse and dplyr to clean and prepare the data.

- **Algorithm Selection:** Choose algorithms available in R packages such as caret, randomForest, nnet and many others.

- **Model Training:** Train models using R functions like train() from the caret package or specific model functions like lm(), glm(), or rpart().

- **Prediction:** Make predictions using predict() functions on the trained models.

- **Evaluation:** Evaluate model performance using metrics provided by packages like caret, yardstick and visualization packages like ggplot2.

Applications in Agriculture

- **Yield Prediction:** Forecasting crop yields based on environmental factors, historical data, and management practices.

- **Disease and Pest Detection:** Identifying and classifying plant diseases or pest infestations using image analysis or sensor data.

- **Precision Agriculture:** Optimizing resource allocation (water, fertilizers, pesticides) based on site-specific needs.

- **Crop Management:** Informing decisions on planting times, irrigation schedules, and harvesting strategies.

- **Soil Analysis:** Classifying soil types, predicting nutrient levels, and identifying areas requiring specific amendments.

Types of machine learning

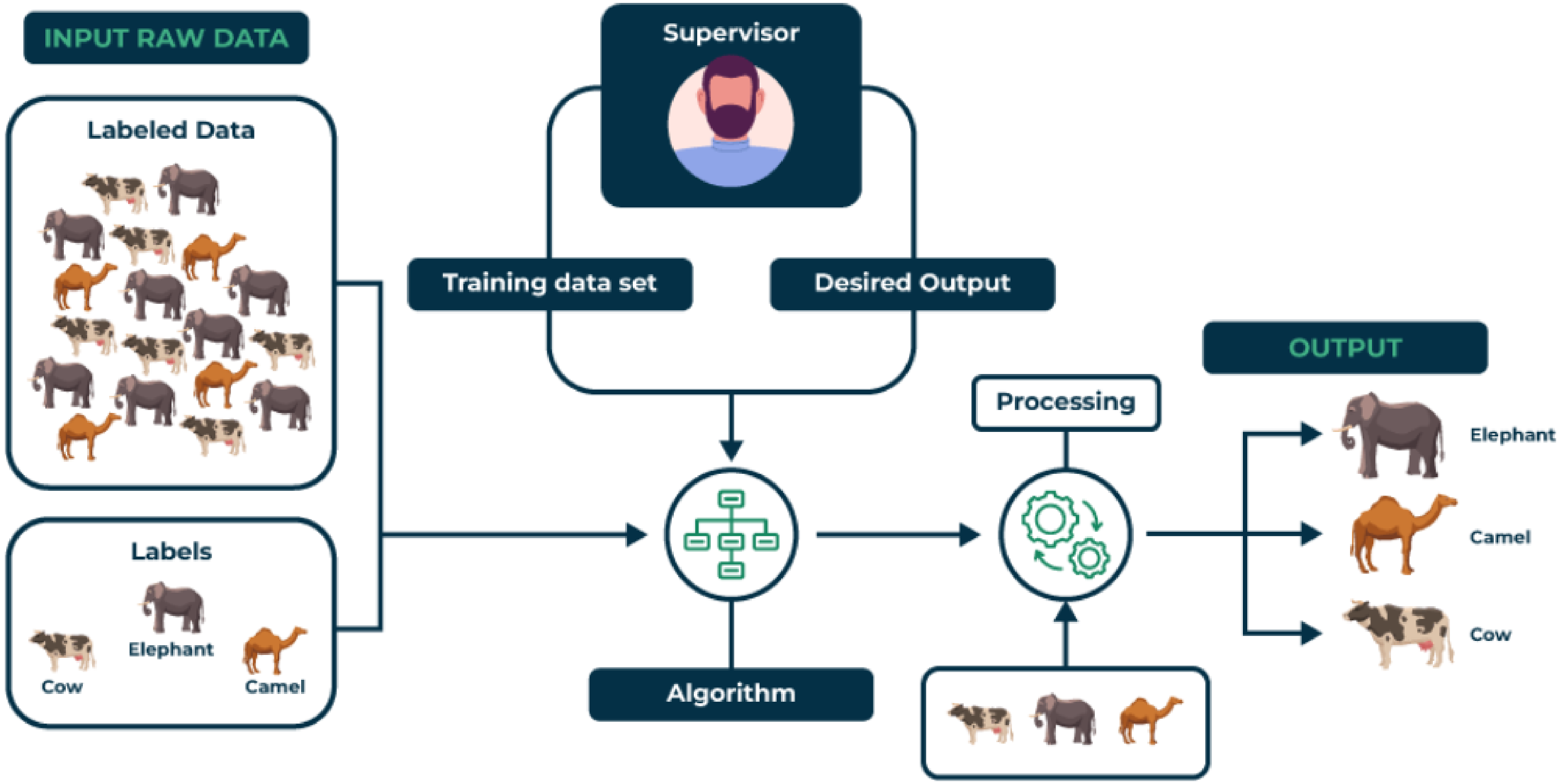
Supervised learning: trains a model using labeled data where each input has a known correct output. The model learns by comparing its predictions with these correct answers and improves over time. e.g regression

Unsupervised learning: This works with unlabeled data where no correct answers or categories are provided. The model's job is to find the data, hidden patterns, similarities or groups on its own. This is useful in scenarios where labeling data is difficult or impossible .e.g clustering

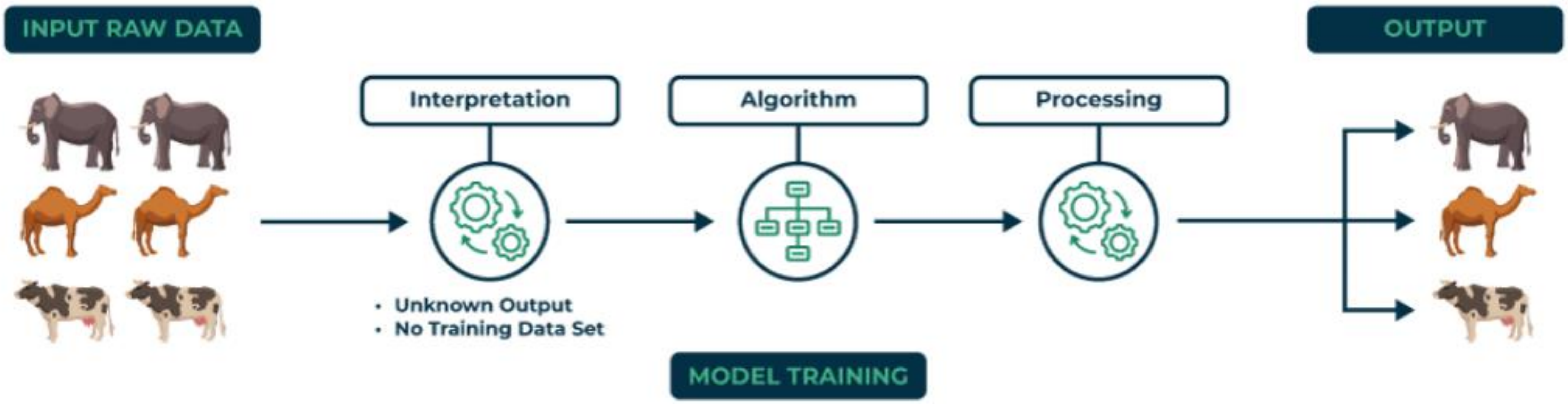
Reinforcement Learning

trains an agent to make decisions by interacting with an environment. Instead of being told the correct answers, agent learns by trial and error method and gets rewards for good actions and penalties for bad ones.. This approach is good for problems having sequential decision making such as robotics, gaming and autonomous systems.

Supervised learning



Unsupervised learning



Algorithms in machine learning

- Algorithms are methods and models used by machines to learn from existing data.
- Examples:

- **Classification:** Predicting a categorical output (e.g., spam or not spam, disease presence or absence). Examples include Logistic Regression, Support Vector Machines (SVMs), Decision Trees, and Neural Networks.

- **Regression:** Predicting a continuous numerical output (e.g., house prices, temperature). Examples include Linear Regression, Ridge Regression, and Lasso Regression.

Algorithms in machine learning

- **Clustering:** Grouping similar data points together (e.g., customer segmentation). Examples include K-Means, Hierarchical Clustering.

- **Dimensionality Reduction:** Reducing the number of variables in a dataset while retaining important information (e.g., for visualization or noise reduction). Examples include Principal Component Analysis (PCA).