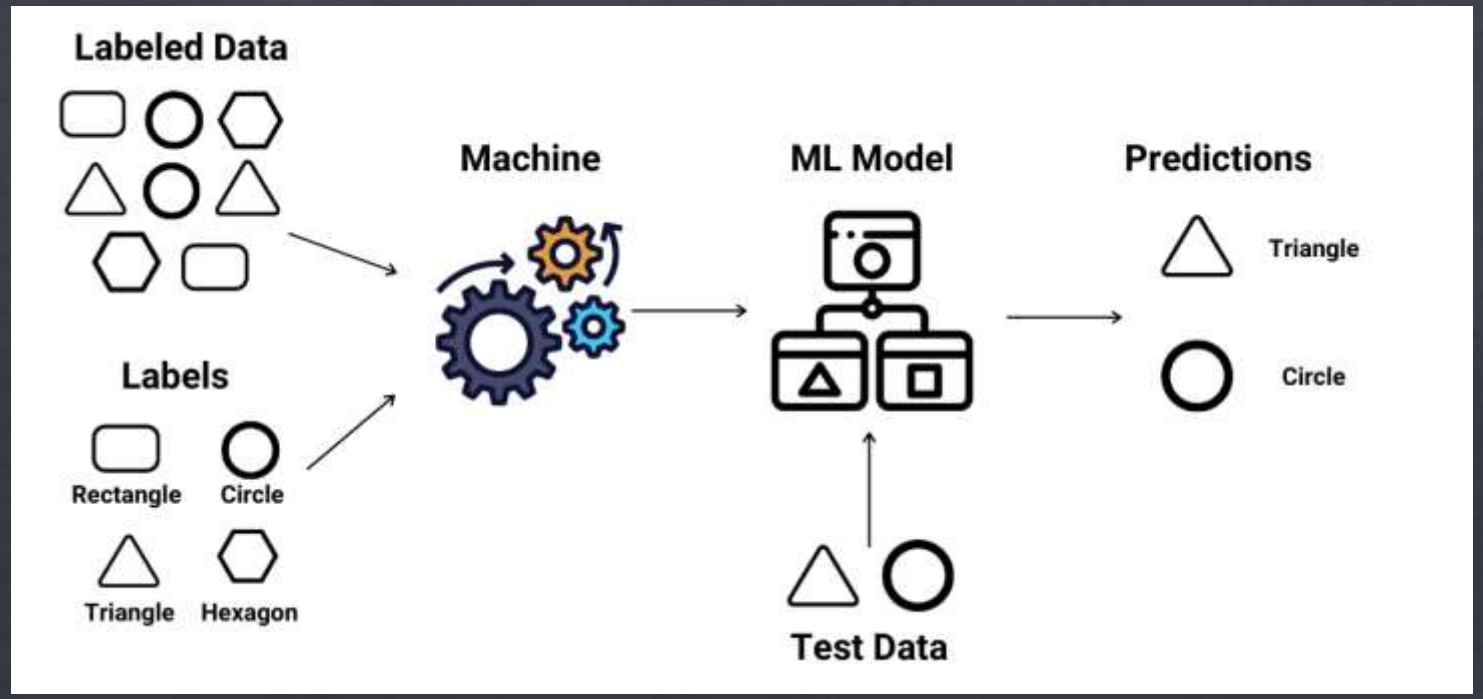
Supervised Machine Learning

Supervised Machine Learning is a type of Machine Learning in which machines are trained using well “labeled” training data and on that data, the machine predicts the output.

The labeled data means some input data is already tagged with the correct output.



Source: enjoy algorithms

Google Developer Student Clubs
Jamia Hamdard

Steps to implement Supervised Machine Learning Algorithms:

- First Determine the type of training dataset.
- Collect / Gather the labeled training data.
- Split the dataset into training, testing, and validation dataset.
- Determine the features best fit to predict the output using Feature Engineering Techniques.
- Determine the Suitable algorithms for the model for the dataset.

Types of Classification Algorithms:

- Support Vector Machine (SVM).
- Logistic Regression.
- Random Forest Classification.
- K-Nearest-Neighbors (KNN).
- Naïve Bayes Classifier.

Types of Regression Algorithms:

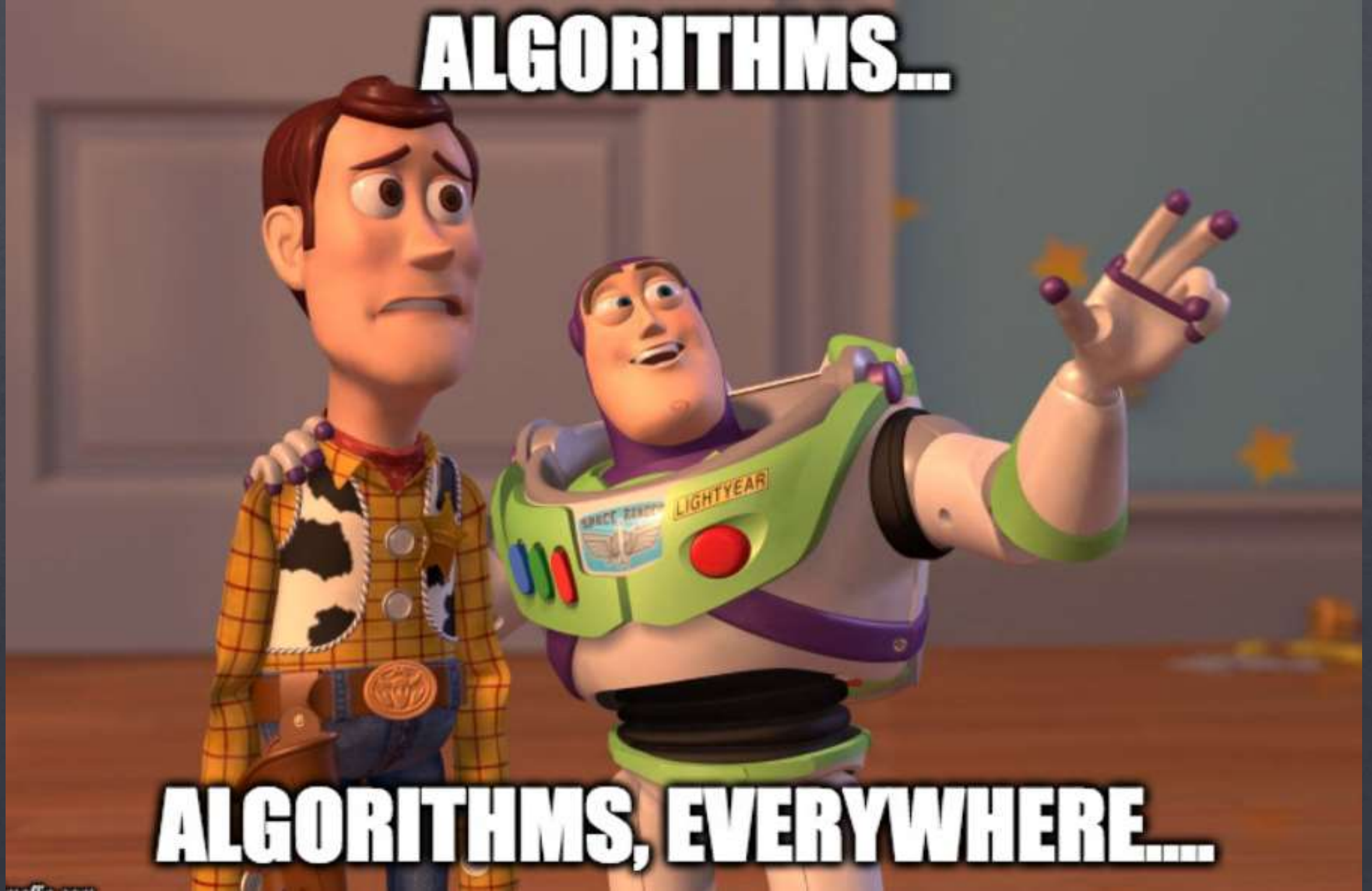
- Linear Regression
- Lasso / Ridge Regression.
- Simple / Multiple Linear Regression.

Advantages:

- The model can predict the output on the basis of prior experience.
- We have an exact idea about the classes of objects.
- Help in real-world problems such as fraud detection, Spam filtering, etc.

Disadvantages:

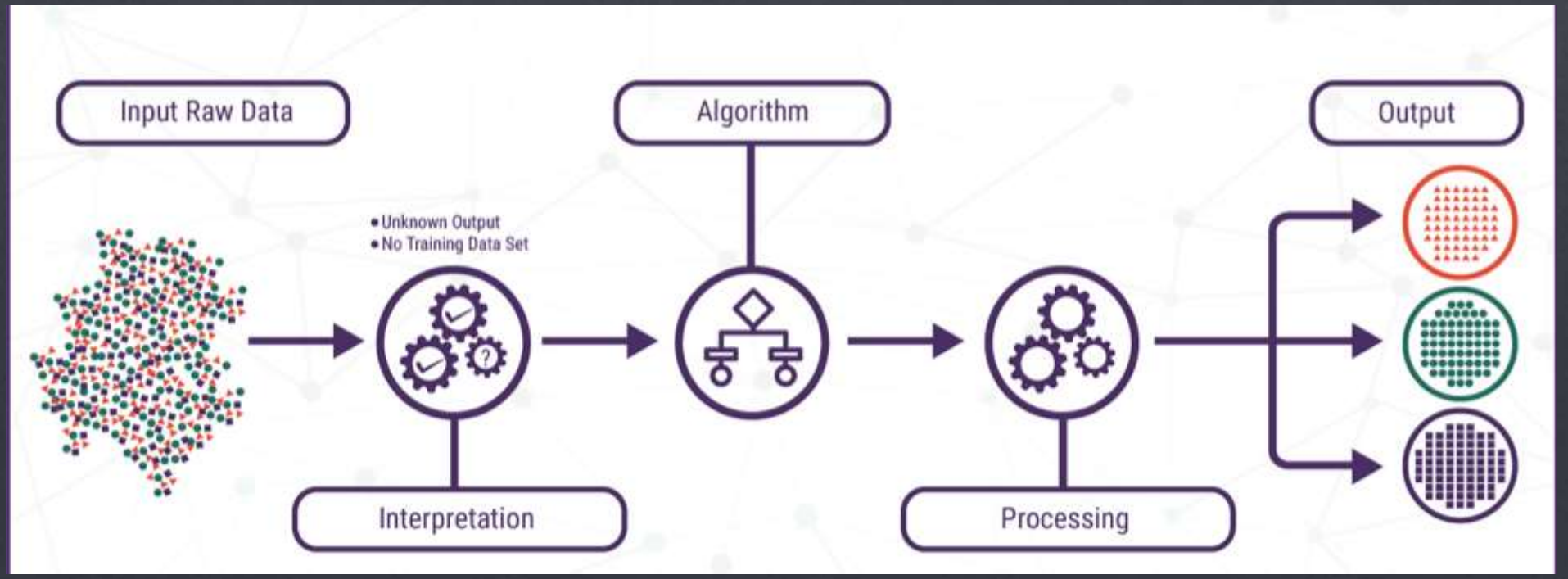
- Not suitable for handling complex tasks.
- It cannot predict the correct output if the test data is different from the training dataset.
- Training requires lots of computation time.


ALGORITHMS...
LIGHTYEAR
ALGORITHMS, EVERYWHERE....

Unsupervised Machine Learning

Unsupervised Machine Learning is a technique in which datasets are unlabeled and it has hidden patterns and insight from the given data.

The goal of Unsupervised learning is to find the underlying structure of the dataset, group the data according to similarities and represent the dataset.



Source: CHI Software

Google Developer Student Clubs
Jamia Hamdard

Steps to Implement Unsupervised Machine Learning Algorithms:

- First Determine the type of training dataset.
- Preparation of data.
- Learn the Underlying Interpretation.
- Determine the features best fit to predict the output using Clustering and Association.
- Determine the Suitable algorithms for the model for the dataset.

Clustering: The method of identifying similar groups of data in a data set. It allows you to automatically split the data into groups according to their similarity. In simple terms, grouping data is based on similarities.

Techniques:

- K-Means Clustering.
- Hierarchical Clustering.
- Principal Component Analysis.
- Neural Networks.

Advantages:

- It is used for more complex tasks as compared to Supervised Machine Learning.
- It can see what human minds cannot visualize.
- It is preferable as it is easier to get unlabeled data than labeled data.

Disadvantages:

- Less Accuracy in the prediction as input data is not known and not in classes of objects.
- Results may be unpredictable and difficult to understand.
- It might right human intervention to understand the patterns.