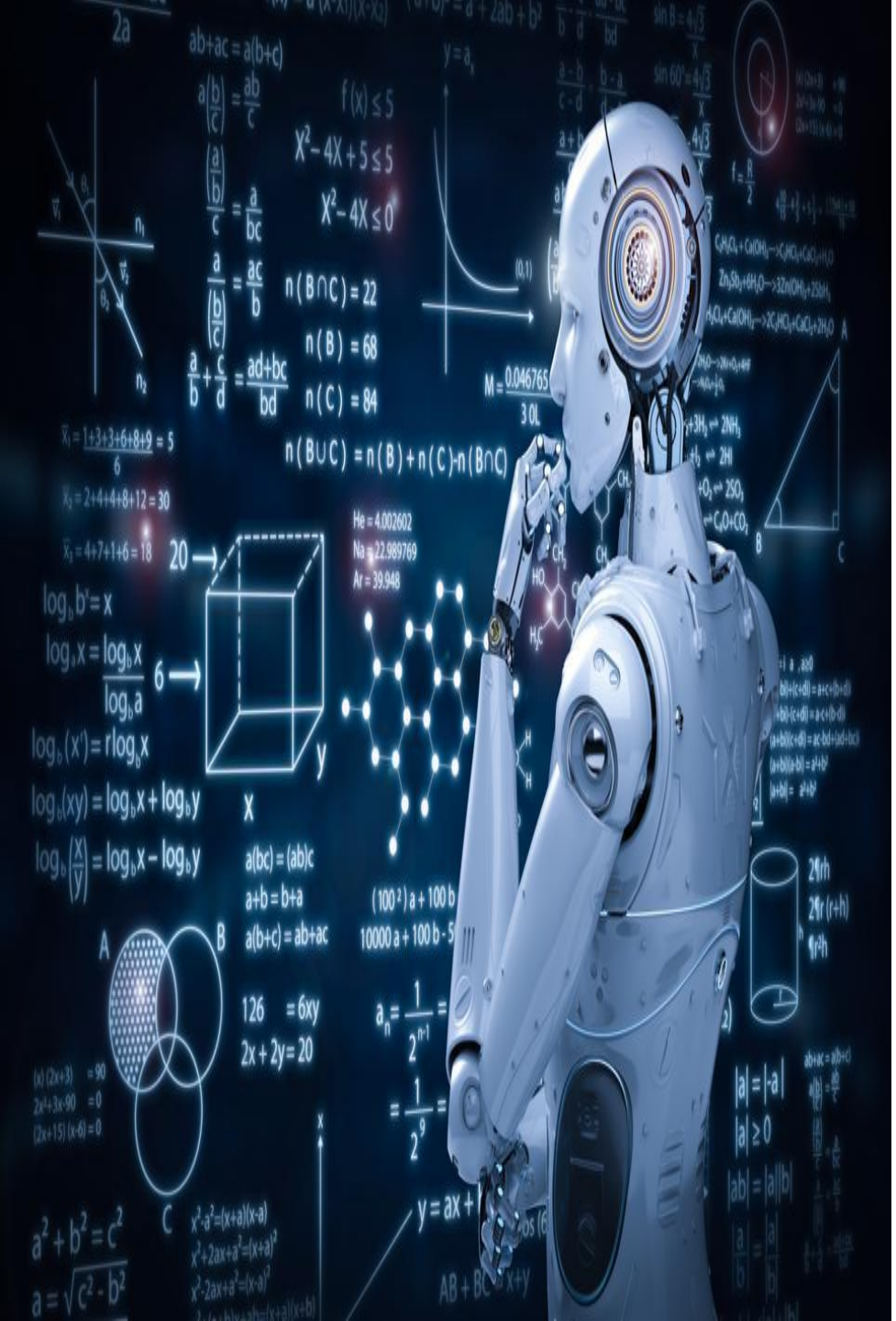


Practical Machine Learning

Day 2: Mar24 DBDA

Kiran Waghmare

Agenda

- What is machine learning?
- Algorithm types of Machine learning
- Supervised and Unsupervised Learning
- Uses of Machine learning
- Evaluating ML techniques
- Introduction to Scikit Learn

Artificial Intelligence

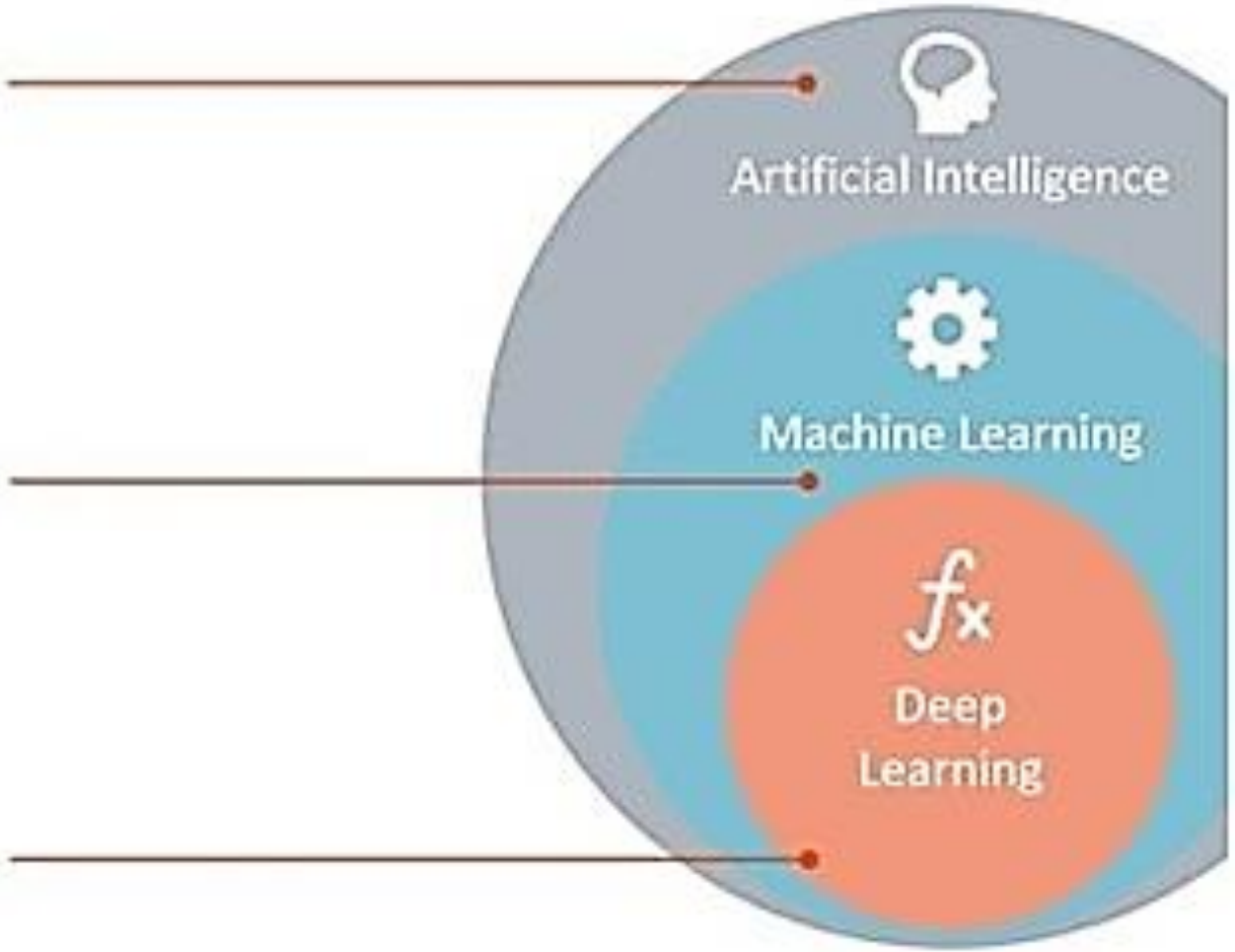
Any technique which enables computers to mimic human behavior.

Machine Learning

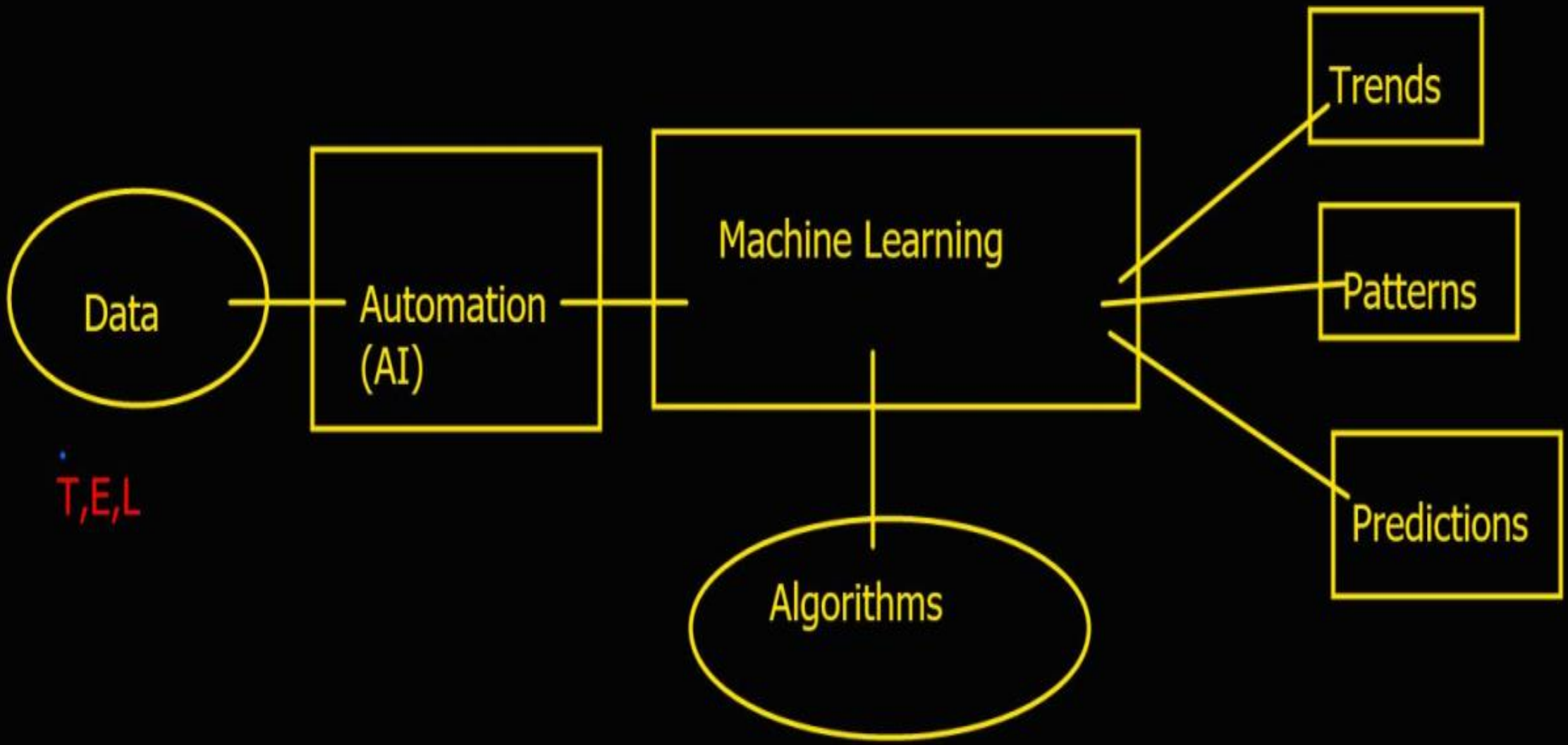
Subset of AI techniques which use statistical methods to enable machines to improve with experiences.

Deep Learning

Subset of ML which make the computation of multi-layer neural networks feasible.



Machine Learning Modelling



T,E,L

What is Machine Learning Model?

- **Definition:**
  - Machine Learning is a concept which allows the machine
  - **to learn from examples and experience**,
  - and that **too without being explicitly programmed**.

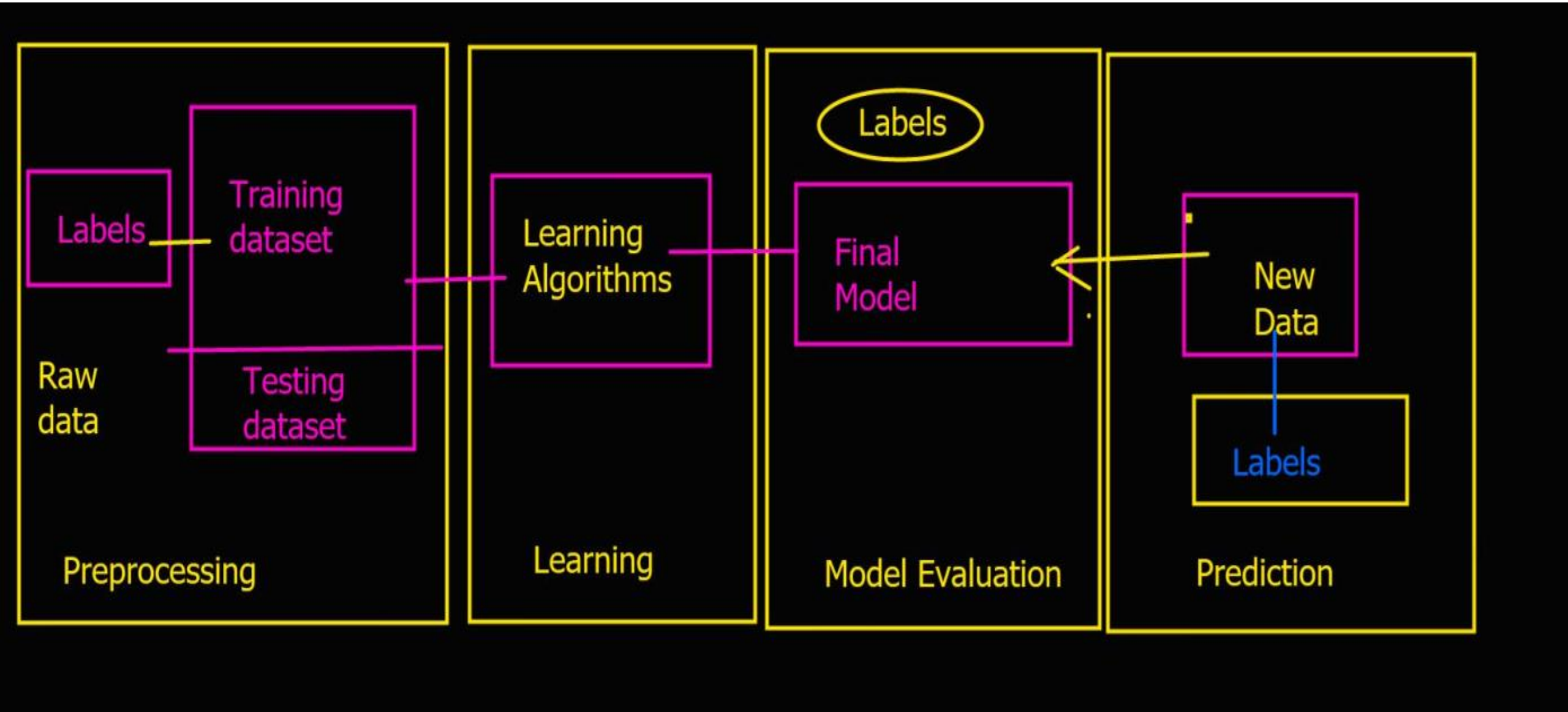
- Machine Learning algorithms are an evolution of **normal algorithms**.
- They make your **programs "smarter",** by **allowing them to automatically learn** from the data you provide.
- The algorithm is mainly divided into:
  - **Training Phase**
  - **Testing phase**

Features of Machine Learning:

- Machine learning **uses data to detect various patterns** in a given dataset.
- It can **learn from past data** and improve automatically.
- It is a **data-driven technology**.
- Machine learning is much **similar to data mining** as it also deals with a huge amount of data.

- Following are some key points that show the importance of Machine Learning:
  - **Rapid increment** in the production of data
  - **Solving complex problems**, which are difficult for a human
  - **Decision-making in various sectors** including finance
  - Finding **hidden patterns and extracting useful information** from data.

ML Lifecycle



4. Steps in Python Machine Learning

- We follow the following steps in Machine Learning Using Python-
- Collecting data.
- Filtering data.
- Analyzing data.
- Training algorithms.
- Testing algorithms.
- Using algorithms for future predictions.

Training Phase

- You take a randomly selected specimen of mangoes from the market (**training data**),
- make a table of all the **physical characteristics** of each mango,
  - like **color, size, shape, grown** in which part of the country,
  - **sold by which vendor**, etc (**features**),
  - along with the **sweetness, juiciness, ripeness** of that mango (**output variables**).
- You feed this data to the machine learning algorithm (**classification/regression**), and it learns a model of the correlation between an average mango's physical characteristics, and its quality.

Testing Phase

- Next time when you go shopping, you will **measure the characteristics of the mangoes** which you are purchasing(**test data**)and **feed it to the Machine Learning algorithm**.
- It will **use the model** which was **computed earlier to predict**
  - if the mangoes are sweet, ripe and/or juicy.
- The algorithm may internally **use the rules**, similar to the one you manually wrote earlier (for eg, a **decision tree**).
- Finally, you can now shop for mangoes with great confidence, without worrying about the details of how to choose the best mangoes.

Conclusion as an Algorithm

- You know what! you can make your algorithm **improve over time** (**reinforcement learning**) so that it **will improve its accuracy** as it gets trained on more and more training dataset.
- In case it makes a wrong prediction it will update its rule by itself.
- The best part of this is, you can use the **same algorithm to train different models.**
- You can create one each for predicting the quality of apples, grapes, bananas, or whatever you want.

Challenges and Limitations of Machine Learning

- The primary challenge of machine learning is the lack of data or the **diversity in the dataset.**
- A machine cannot learn if there **is no data available**.
- Besides, a dataset with a lack of diversity gives the machine a hard time.
- A machine **needs to have heterogeneity** to learn meaningful insight.
- It is rare that an algorithm can extract information when there are no or few variations.
- It is recommended to have at least 20 observations per group to help the machine learn.
- This constraint leads to poor evaluation and prediction.