

INTRODUCTION TO MACHINE LEARNING

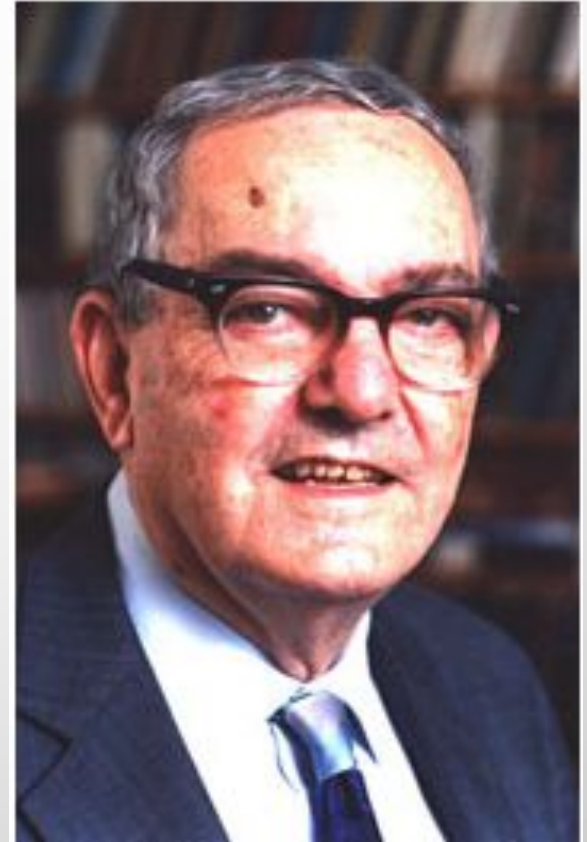


Machine Learning

- Herbert Alexander Simon:

Learning is any process by which a system improves performance from experience.

- Machine Learning is concerned with computer programs that automatically improve their performance through experience.



Herbert Simon

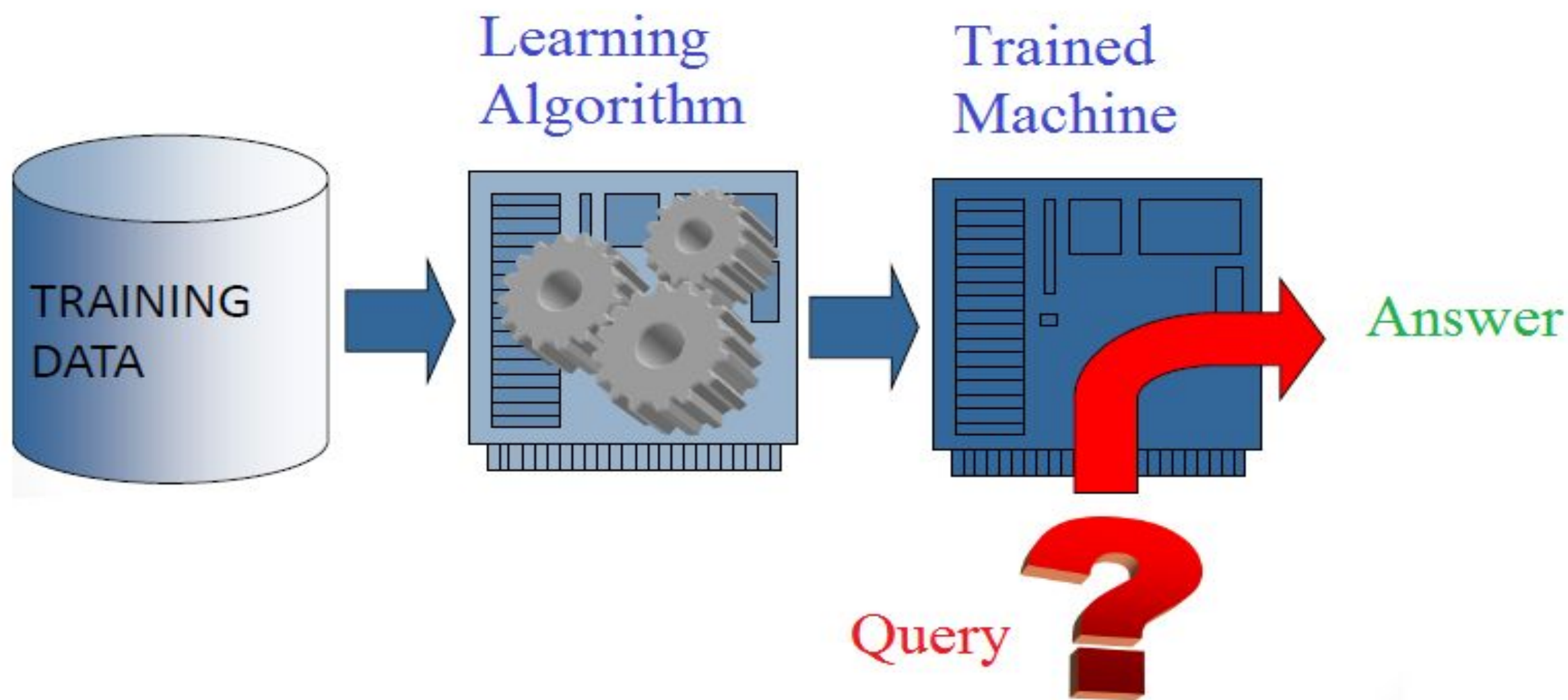
The Learning



Why Machine Learning?

- Develop systems that can automatically adapt and customize themselves to individual users.
 - Personalized news or mail filter
- Discover new knowledge from large databases (data mining).
 - Market basket analysis (e.g. diapers and Fruits)
- Ability to mimic human and replace certain monotonous tasks which require some intelligence.
 - like recognizing handwritten characters
- Develop systems that are too difficult/expensive to construct manually because they require specific detailed skills or knowledge tuned to a specific task (knowledge engineering bottleneck)

Machine Learning



Example: Classification using ML

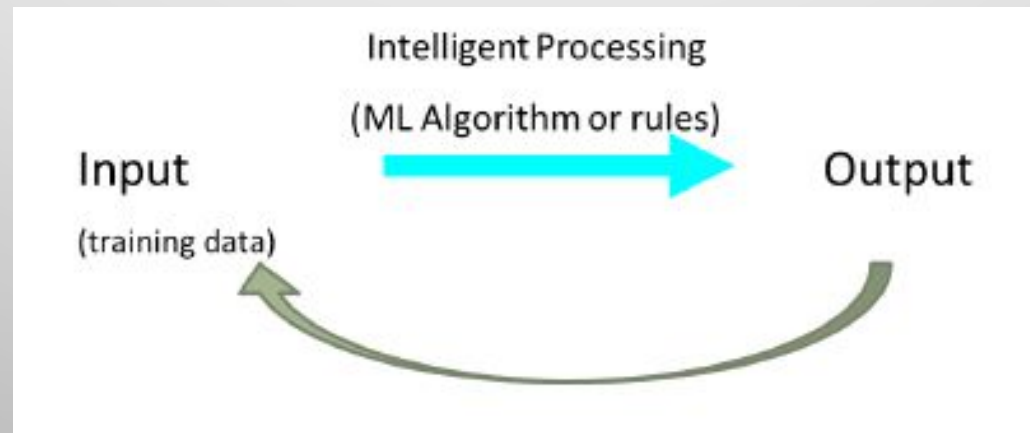
Image processing:

- ❑ Machine learning can be used in classification Images & objects in an image
 - ❖ Ship
 - ❖ Water
 - ❖ Rock
 - ❖ Iron object
 - ❖ Fiber Object etc.,

Does this really help?

The main advantage of ML

- ❑ Learning and writing an algorithm
 - ❑ Its easy for human brain but it is tough for a machine. It takes some time and good amount of training data for machine to accurately classify objects
- ❑ Implementation and automation
 - ❑ This is easy for a machine. Once learnt a machine can process one million images without any fatigue where as human brain can't
- ❑ That's why ML with big data is a deadly combination



Applications of Machine Learning

1. Banking / Telecom / Retail

☐ Identify:

- ☐ Prospective customers
- ☐ Dissatisfied customers
- ☐ Good customers
- ☐ Bad payers

☐ Obtain:

- ☐ More effective advertising
- ☐ Less credit risk
- ☐ Fewer fraud
- ☐ Decreased churn rate



2. Biomedical / Biometrics

❑ Medicine:

- ❑ Screening
- ❑ Diagnosis and prognosis
- ❑ Drug discovery

❑ Security:

- ❑ Face recognition
- ❑ Signature / fingerprint / iris verification
- ❑ DNA fingerprinting



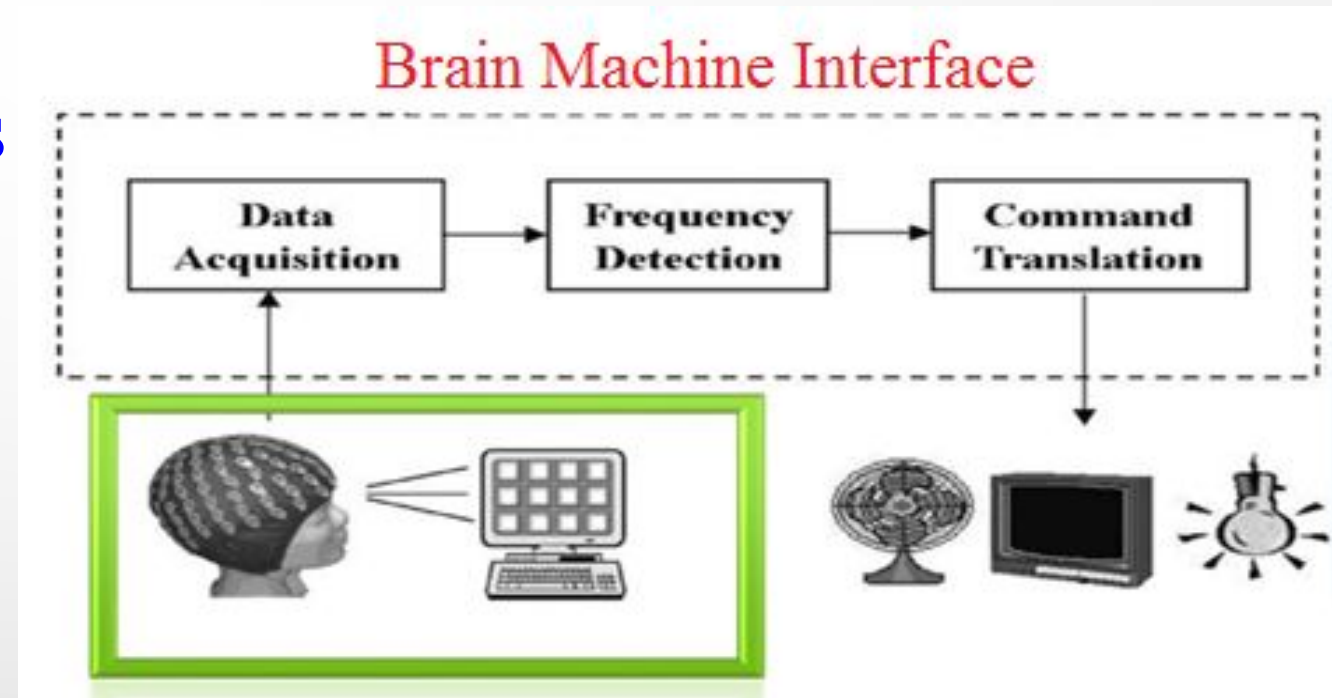
3. Computer / Internet

❑ Computer interfaces:

- ❑ Troubleshooting wizards
- ❑ Handwriting and speech
- ❑ Brain waves

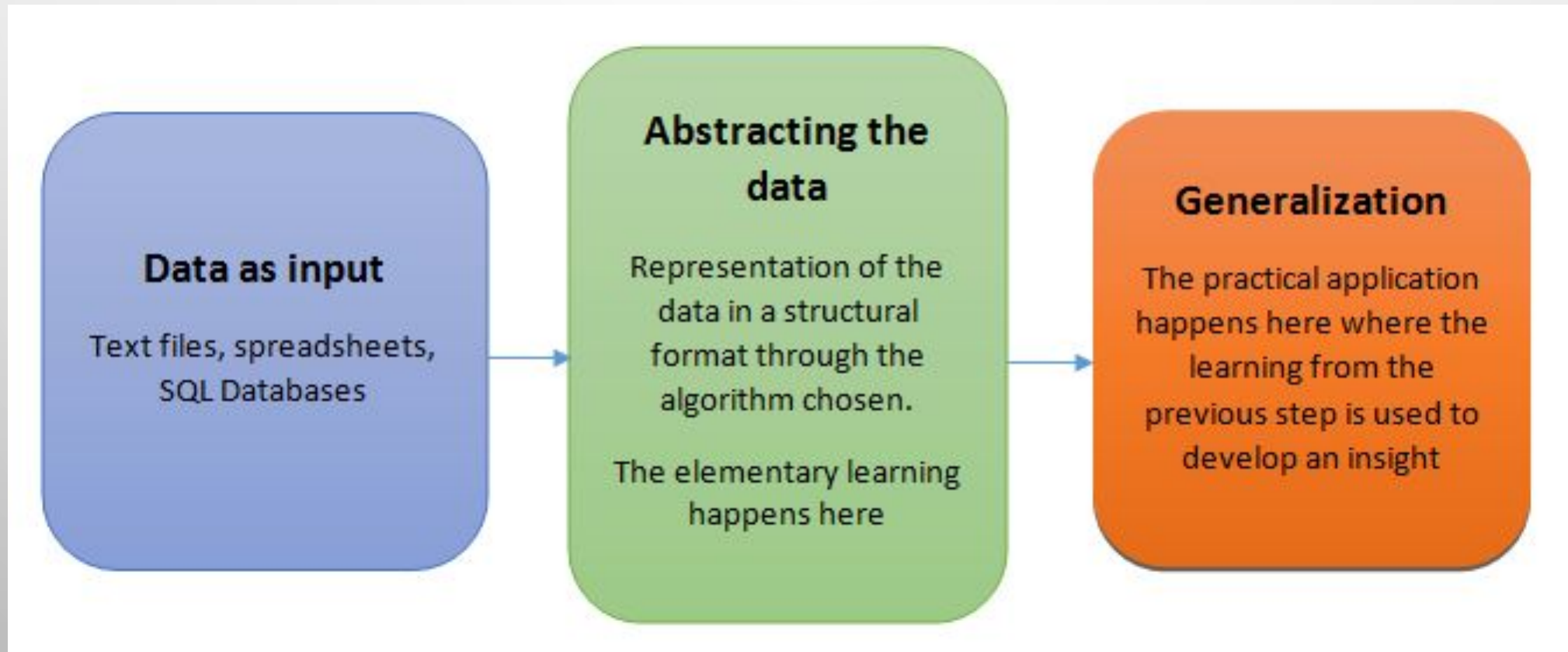
❑ Internet

- ❑ Hit ranking
- ❑ Spam filtering
- ❑ Text categorization
- ❑ Text translation
- ❑ Recommendation



How exactly do we teach machines?

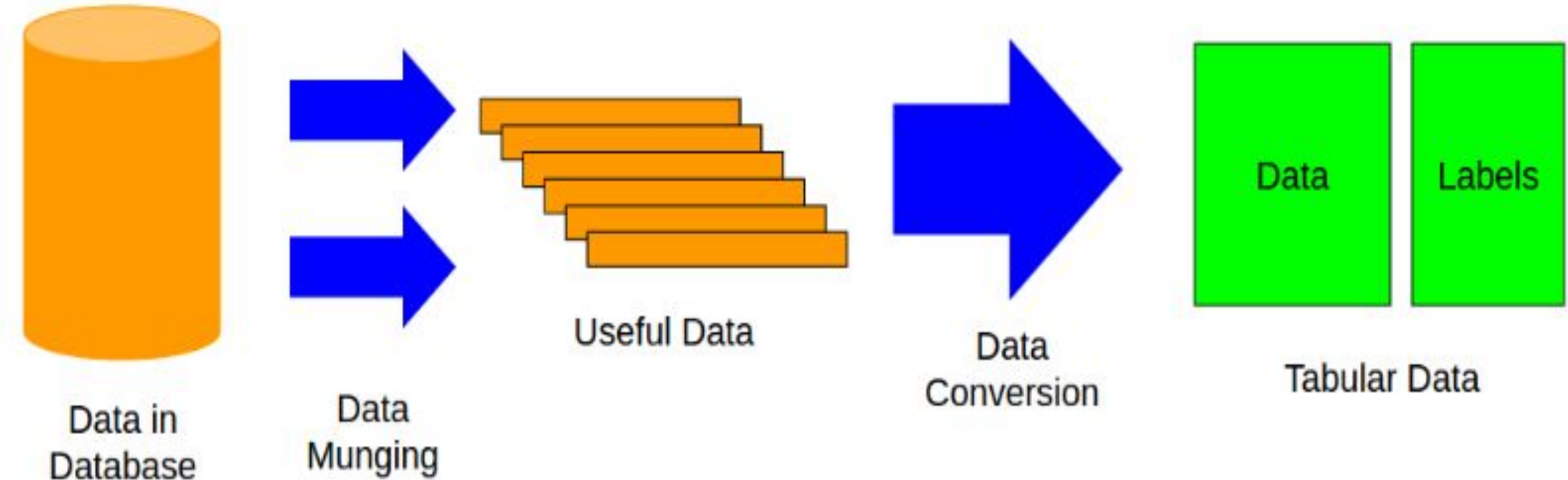
Teaching the machines involve a structural process where every stage builds a better version of the machine. For simplification purpose, the process of teaching machines can broken down into 3 parts:



Basic steps used in Machine Learning

There are 5 basic steps used to perform a machine learning task:

- 1. Collecting data:** Be it the raw data from excel, access, text files etc., this step (gathering past data) forms the foundation of the future learning. The better the variety, density and volume of relevant data, better the learning prospects for the machine becomes



3. Training a model:

- This step involves choosing the appropriate algorithm and representation of data in the form of the model.
- The cleaned data is split into two parts – train and test (proportion depending on the prerequisites);
- The first part (training data) is used for developing the model. The second part (test data), is used as a reference.

3. Evaluating the model:

- To test the accuracy, the second part of the data (holdout / test data) is used.
- This step determines the precision in the choice of the algorithm based on the outcome.
- A better test to check accuracy of model is to see its performance on data which was not used at all during model build.

3. Improving the performance:

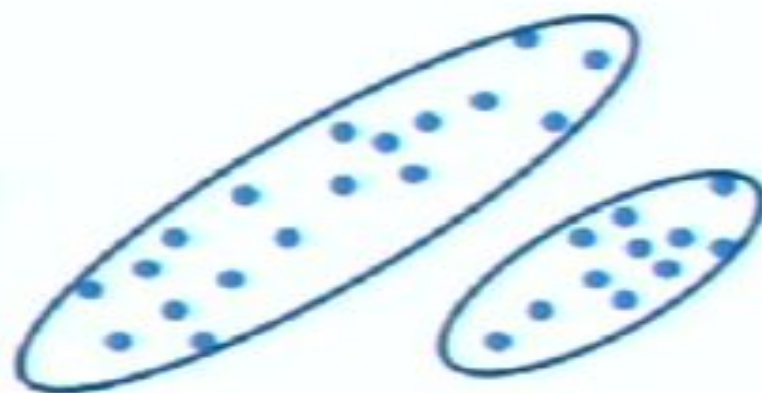
- This step might involve choosing a different model altogether or introducing more variables to augment the efficiency.
- That's why significant amount of time needs to be spent **in data collection and preparation.**

What Machine Learning consists of

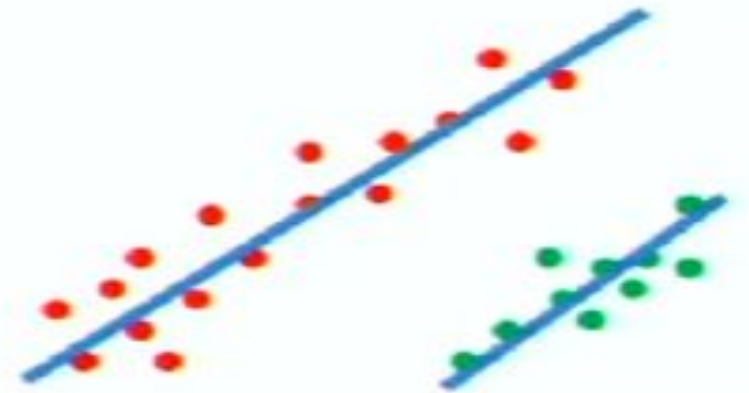
Classification



Clustering

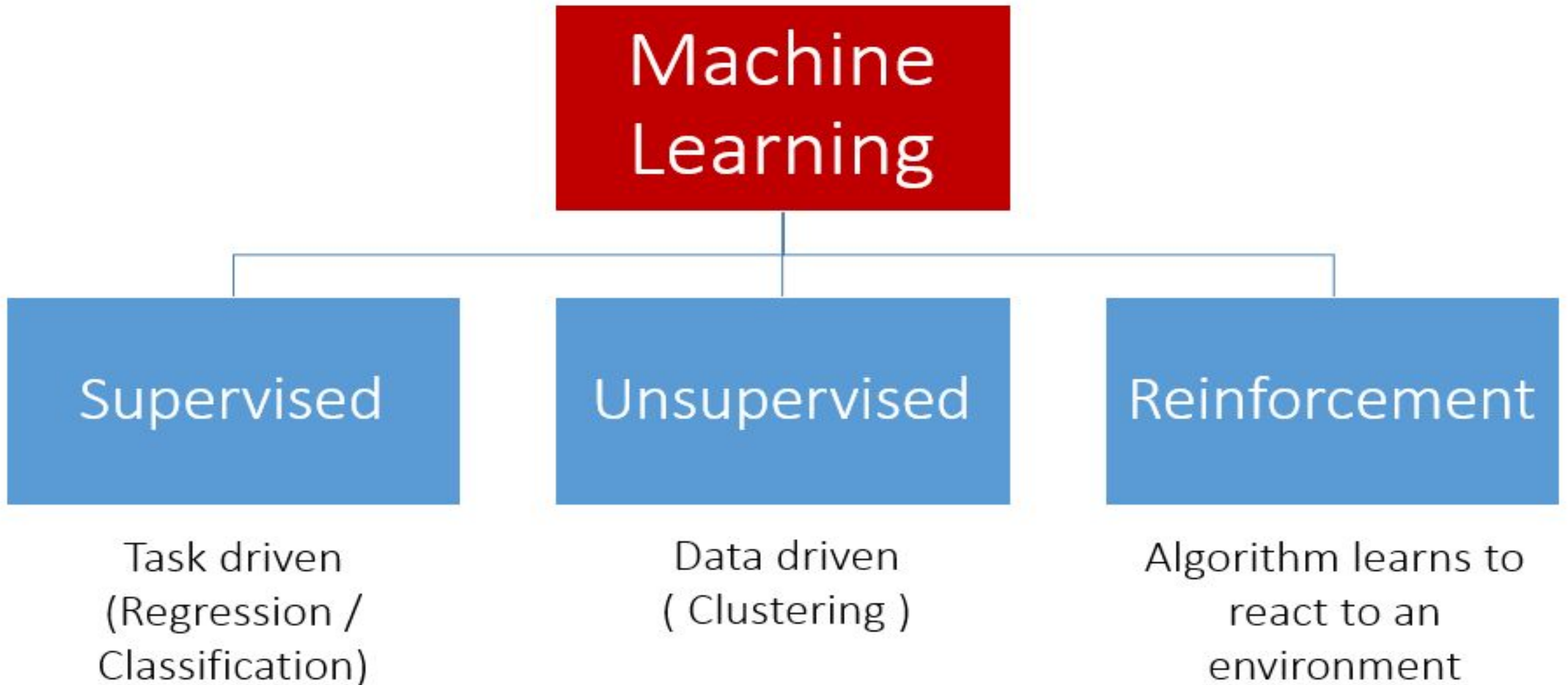


Regression



- Classification and clustering look similar
 - But the technology is quite different
- Classification and regression look quite different
 - But the technology is fairly similar

The Types of Machine Learning Algorithms



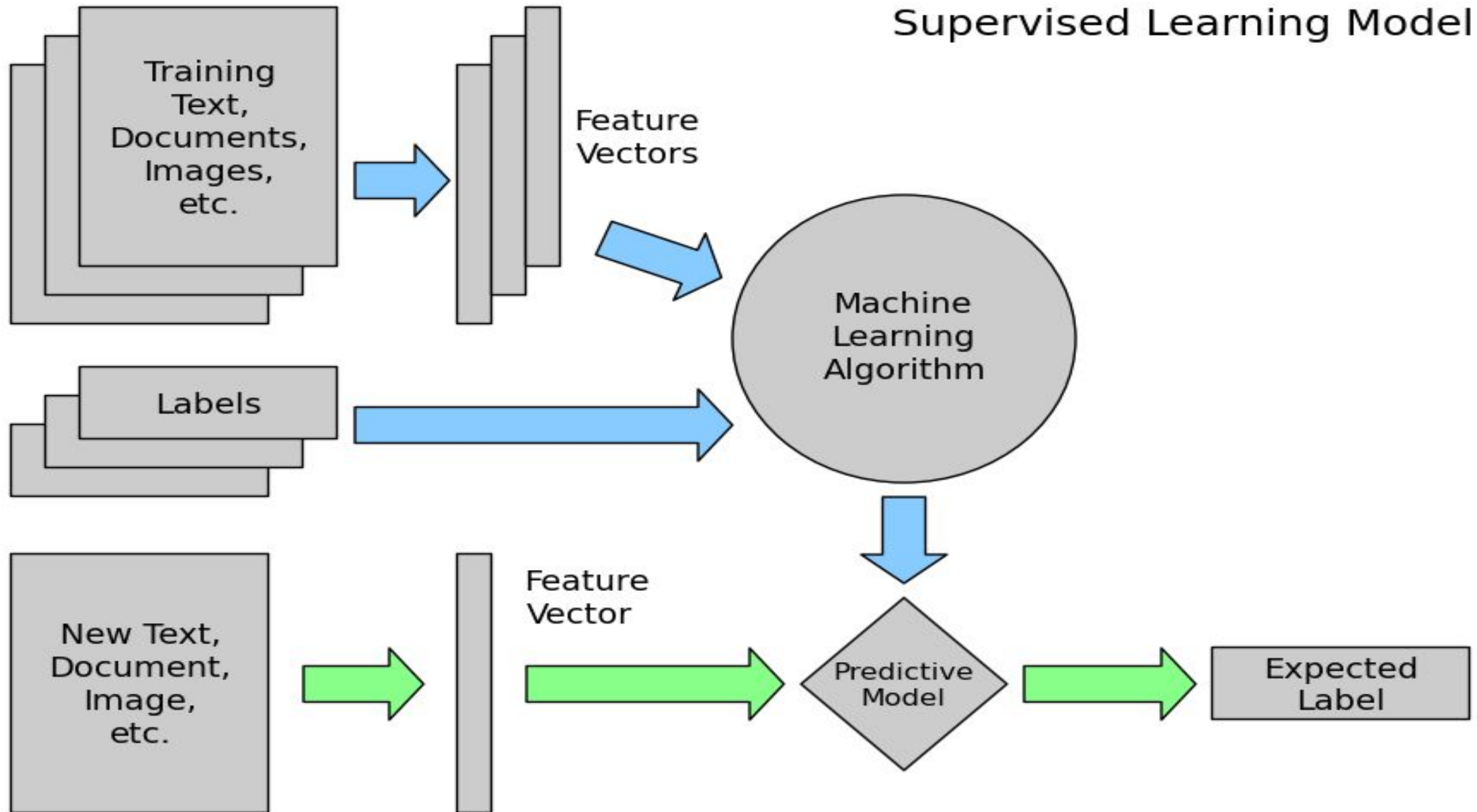
Supervised Learning / Predictive models:

- Predictive model as the name suggests is used to predict the future outcome based on the historical data.
- Predictive models are normally given clear instructions right from the beginning as in what needs to be learnt and how it needs to be learnt.

For example:

Supervised Learning is used when a marketing company is trying to find out which customers are likely to churn. We can also use it to predict the likelihood of occurrence of perils like earthquakes, tornadoes etc. with an aim to determine the Total Insurance Value. Some examples of algorithms used are: Nearest neighbour, Naïve Bayes, Decision Trees, Regression etc.

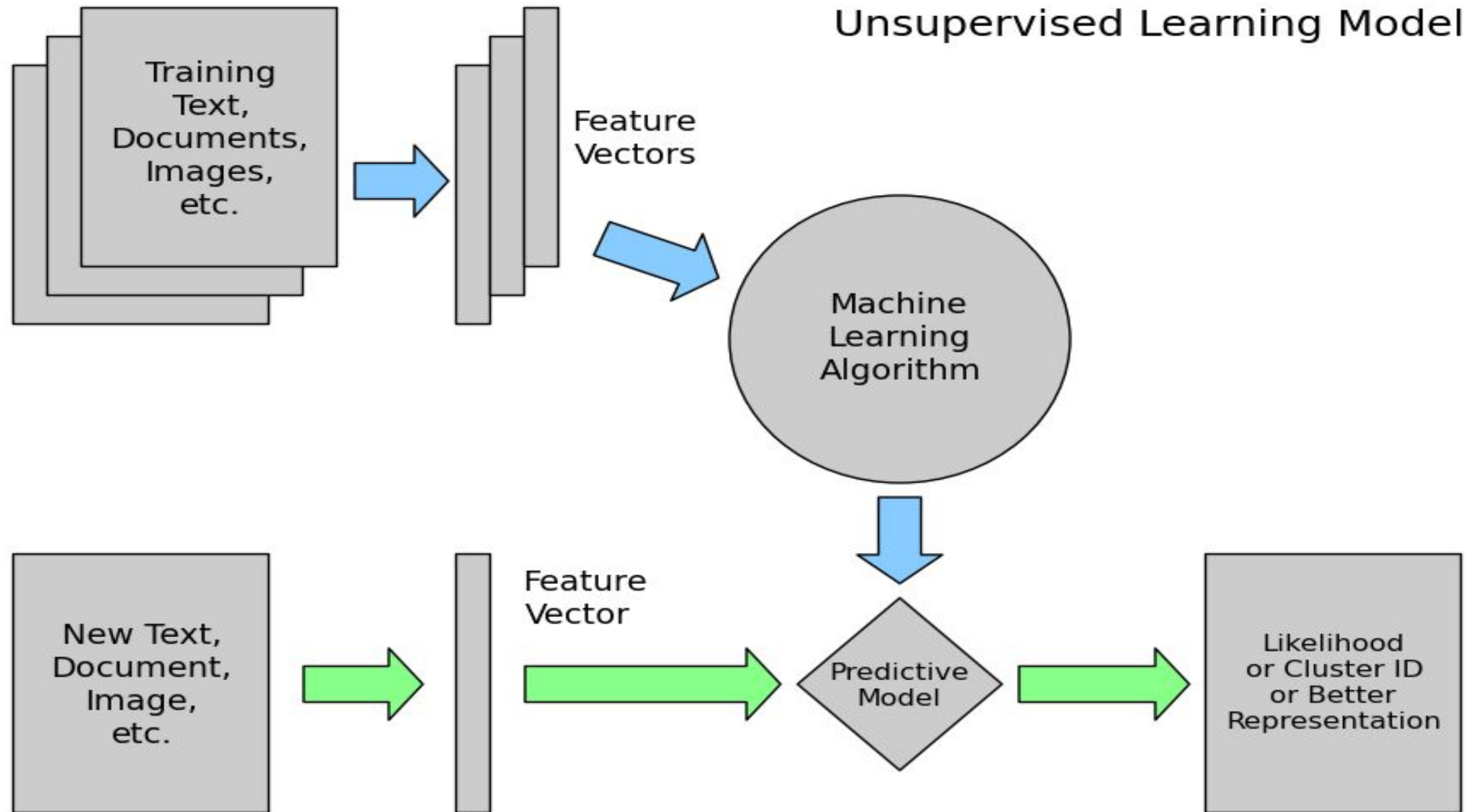
Supervised Learning Model



Unsupervised learning / Descriptive models:

- It is used to train descriptive models where no target is set and no single feature is important than the other.
- The case of unsupervised learning can be: When a retailer wishes to find out what are the combination of products, customers tends to buy more frequently.
- Furthermore, in pharmaceutical industry, unsupervised learning may be used to predict which diseases are likely to occur along with diabetes.
- Example of algorithm used here is: K- means Clustering Algorithm

Unsupervised Learning Model



RESOURCES: DATASETS

- UCI REPOSITORY: [HTTP://WWW.ICS.UCI.EDU/~MLEARN/MLREPOSITORY.HTML](http://www.ics.uci.edu/~mllearn/mlrepository.html)
- UCI KDD ARCHIVE:
[HTTP://KDD.ICS.UCI.EDU/SUMMARY.DATA.APPLICATION.HTML](http://kdd.ics.uci.edu/summary.data.application.html)
- STATLIB: [HTTP://LIB.STAT.CMU.EDU/](http://lib.stat.cmu.edu/)
- DELVE: [HTTP://WWW.CS.UTORONTO.CA/~DELVE/](http://www.cs.utoronto.ca/~delve/)

THE END