

# Introduction to Machine Learning

BY:

**DR. VIPIN KUMAR**

DEPARTMENT OF COMPUTER SCIENCE & IT  
MAHATMA GANDHI CENTRAL UNIVERSITY  
MOTIHARI, BIHAR



# What is Machine Learning...???

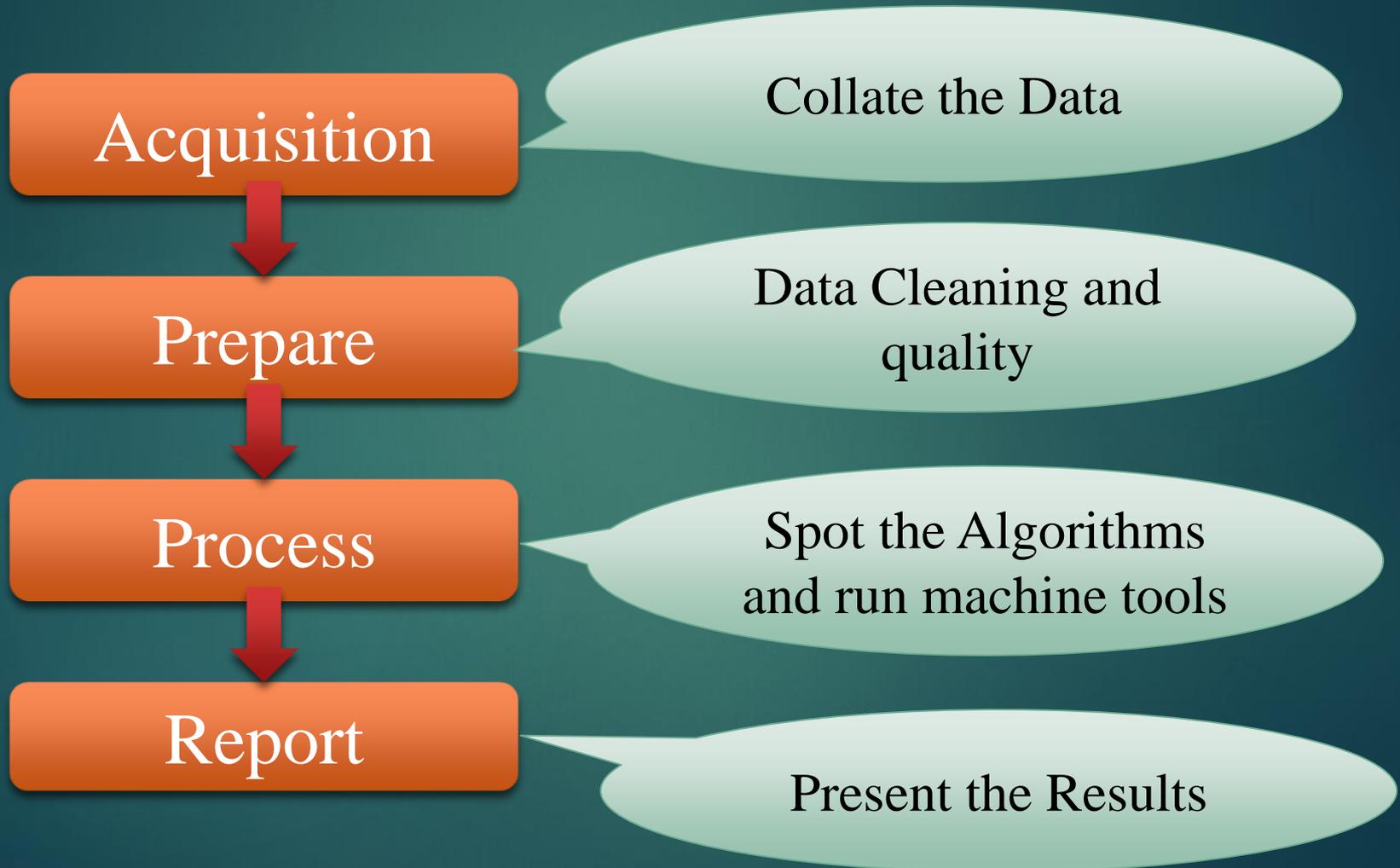
## Definitions:

- ▶ Arthur Samuel (1959): Machine Learning is Field of study that gives computers the ability to learn without being explicitly programmed.
- ▶ When a deterministic algorithms fails to solve the real life problem, then machine learning is applied to solve the problem.
- ▶ It obtains a hypothesis from the previously known information to predict future unknown scenario.

# Outlines.....

- ▶ Defining Machine Learning
- ▶ Machine Learning Process
- ▶ Types of Machine Learning
- ▶ Supervised Learning
- ▶ Unsupervised Learning
- ▶ Reinforcement Learning
- ▶ List of dataset online sources
- ▶ List of useful language for machine learning modeling

# Machine Learning Process



# Applications

- ▶ **Large datasets from growth of automation/web.**
  - ▶ E.g., Web click data, medical records, biology, engineering
- ▶ **Applications can't program by hand.**
  - ▶ E.g., Autonomous helicopter, handwriting recognition, most of Natural Language Processing (NLP).
- ▶ **Self-customizing programs**
  - ▶ E.g., Amazon, Netflix product recommendations
- ▶ **Understanding human learning (brain, real AI).**

# Type Machine Learning Algorithms

- ▶ **Supervised Learning**
  - ▶ Classification
  - ▶ Regression
- ▶ **Unsupervised Learning**
  - ▶ Clustering
- ▶ **Reinforcement Learning**

# Types of Supervised Learning

## Classification:

- ▶ It utilizes the labeled data for building the model to predict the discrete labels of unknown test samples.
- ▶ Example:
  - ▶ Financial Institution: Credit Scoring by the banks (Low-risk and High-risk)

## Regression

- ▶ It utilizes the labeled data for building the model to predict the continuous labels of unknown test samples.
- ▶ Example:
  - ▶ Housing price prediction

# Applications of Supervised Learning

- Handwriting recognition
- Bioinformatics
- Information retrieval
- OCR (Optical Character recognition)
- Pattern recognition
- Speech recognition
- Spam detection

# Unsupervised Learning

- ▶ **Unsupervised learning** is a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses.
- ▶ The most common **unsupervised learning** method is *cluster analysis*, which is used for exploratory data analysis to find hidden patterns or grouping in data.

# Applications of Unsupervised Learning

- Social network analysis
- Image analysis
- Summarizing news
- Document summarization
- Marketing
- Land use

# Semi-supervised Learning

- It is a task of supervised learning which utilized unlabeled data for the training.
- It is typically employed where small amount of labeled data with large amount of unlabeled data.

# Application of Semi-supervised Learning

- Text classification
- Document classification
- Image classification
- Hyper-spectral image classification
- Streaming data classification

# Reinforcement Learning

- ▶ It is the problem of getting an agent to act in the world so as to maximize its rewards.
- ▶ It utilizes the reward and penalty to optimize the policy.
- ▶ *Example* : Train computers to do many tasks, such as playing backgammon or chess, scheduling jobs, and controlling robot limbs.

# Some References of Datasets Repositories...

- ▶ [www.KDnuggets.com](http://www.KDnuggets.com)
- ▶ [www.archive.ics.uci.edu/ml/](http://www.archive.ics.uci.edu/ml/)
- ▶ [www.data.gov.uk](http://www.data.gov.uk)
- ▶ [www.kaggle.com](http://www.kaggle.com)
- ▶ [www.Labrosa.ee.columbia.edu/millionsong/](http://www.Labrosa.ee.columbia.edu/millionsong/)

# Languages for Machine Learning Algorithms Implementation

- ▶ **Python**
- ▶ Matlab
- ▶ R
- ▶ Weka
- ▶ Rapidminer

# Bibliography:

- E. Alpaydin, Introduction of Machine Learning, Prentice Hall of India, 2006
- T.M. Mitchell, Machine Learning, McGraw-Hill, 1997
- C.M. Bishop, **Pattern Recognition and Machine Learning**, Springer, 2006.
- R.O. Duda, P.E. Hart and D.G. Stork, **Pattern Classification**, John Wiley, and Sons, 2001.
- Vladimir N. Vapnik, **Statistical Learning Theory**, John Wiley and Sons, 1998.
- Shawe-Taylor J. and Cristianini N., **Introduction to Support Vector Machines**, Cambridge University Press, 2000.



Thank You