# Big Data Analytical Techniques:
## AI, ML and DM Definition

# Data is generating a fast speed

- Every minute on Facebook: 510,000 comments are posted, 293,000 statuses are updated, and 136,000 photos are uploaded.
- 300 hours of vedio are uploaded to YouTube each minute.
- Instagram users upload over 100 million photos and videos overyday. That is 69,444 million posts every minute!
- Over 3.5 Billion Google searches are conducted worldwide each minute of everyday. That is **2 trillion searches per year** worldwide. That is over 40,000 search queries per second!

# Data Science

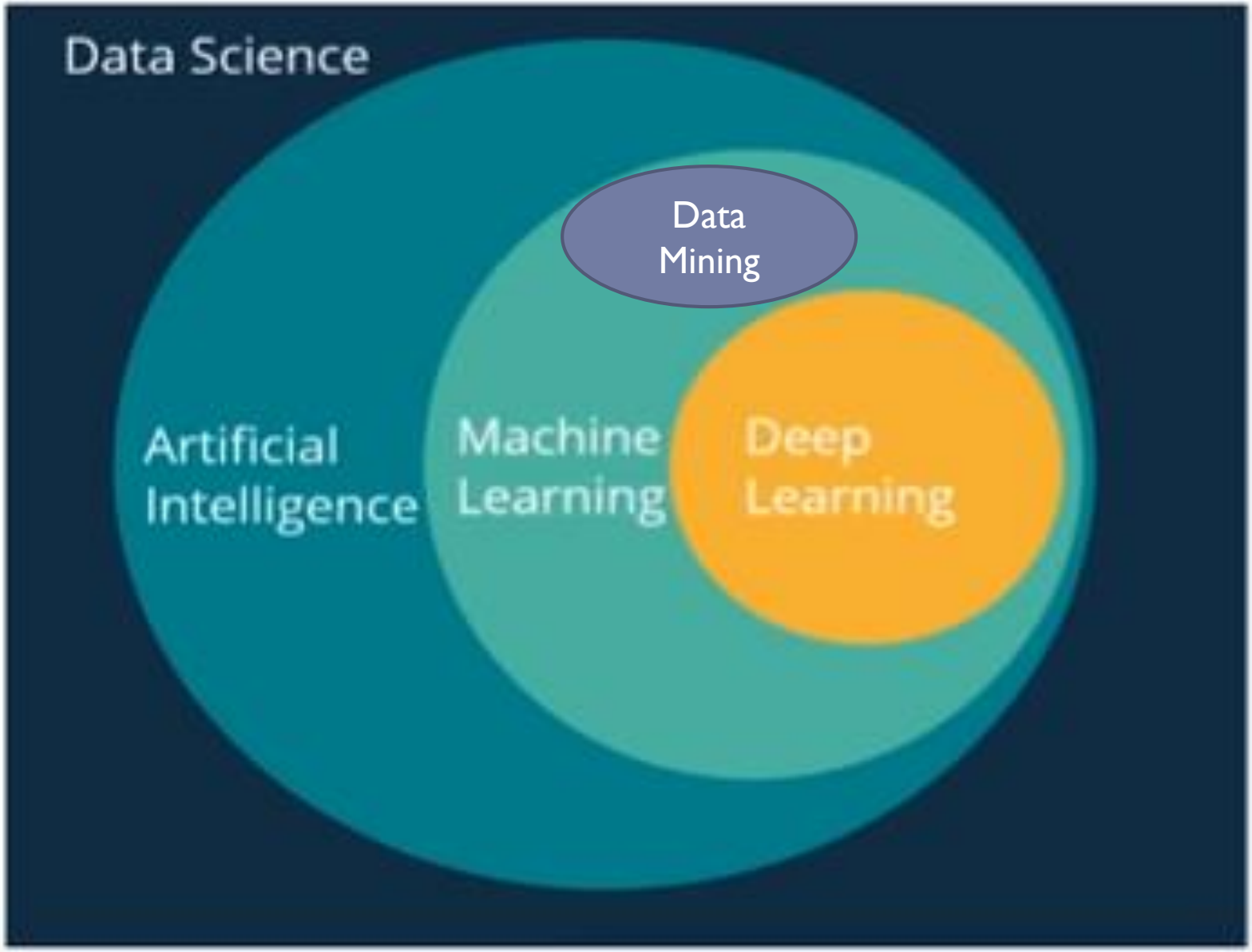
Programming + Statistics + Business

Data Science Domains

- Estimated by 2020, 1.7 MB of data will be created each second by each person over the earth.
- Results into Lots of data!!!!!!!!
- Gigabyte→Terabyte→Petabyte→Exabyte→ →Zettabyte→Yottabyte→…….etc

# What is Data Science?

- Data Science is the process of extracting useful information (knowledge) from data by using different algorithms and techniques.


Data Science
Data
Mining
Artificial
Intelligence
Machine
Learning
Deep
Learning

# Basic Concepts of AI/ML/DM/DL

- Definition of AI:

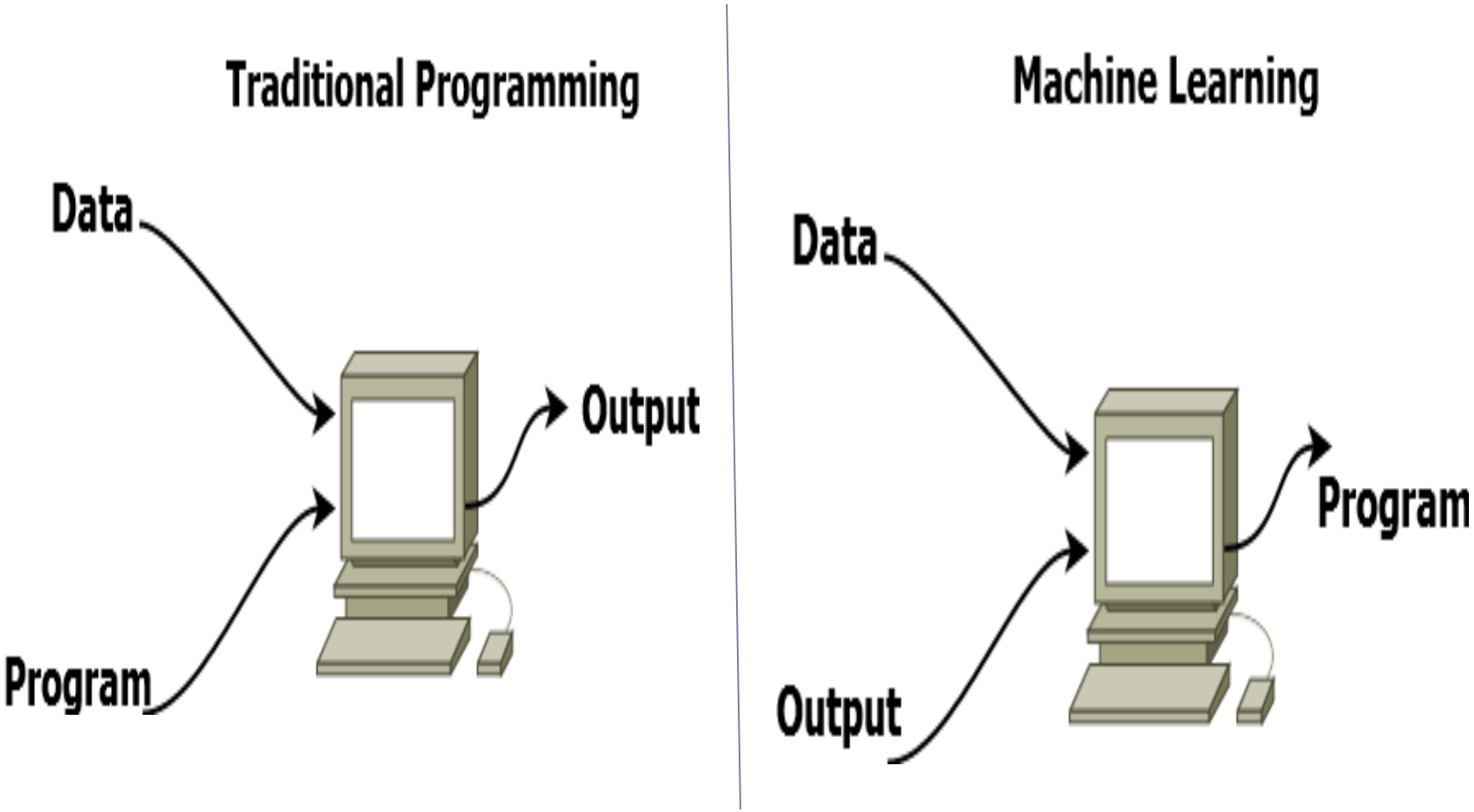
- **Artificial Intelligence is technique which enables machines to mimic human behaviour.**

- **Ex. Robotics, AI-powered Chatbot**

# Definition of ML:

- **Machine learning algorithms build programs based on historic data in order to make predictions or decisions to perform the task without being explicitly programmed.**

# Basic Concepts of AI/ML/DM/DL



# Example: Diabetic Patient Data

```
'preg','plas','pres', 'skin', 'test',
'mass','pedi','age','class
```

```
6,148,72,35,0,33.6,0.627,50,1
1,85,66,29,0,26.6,0.351,31,0
8,183,64,0,0,23.3,0.672,32,1
1,89,66,23,94,28.1,0.167,21,0
0,137,40,35,168,43.1,2.288,33,1
5,116,74,0,0,25.6,0.201,30,0
3,78,50,32,88,31.0,0.248,26,1
10,115,0,0,0,35.3,0.134,29,0
2,197,70,45,543,30.5,0.158,53,1
8,125,96,0,0,0.0,0.232,54,1
4,110,92,0,0,37.6,0.191,30,0
```

# Basic Concepts of AI/ML/DM/DL

**Definition of ML:**

**Machine Learning is a subset of AI technique which uses statistical methods to enables machines to build model for predictions which improve with experiences.**

**(Field of Data Science which does predictive analysis])**

# Basic Concepts of AI/ML/DM/DL

**Definition of DM:**

**Data mining is the process of discovering hidden patterns in large data sets using methods of machine learning, statistics, and database technology.**

# Data Mining

- **Data Mining is the:**
  - Discovery of useful, possibly unexpected, patterns in data
  - Non-trivial extraction of implicit, previously unknown and potentially useful information from data

# Basic Concepts of AI/ML/DM/DL

Definition of Deep Learning (DL):

**Deep Learning is a subset of ML which make use of computation of multi-layer Neural Network.**

# Example

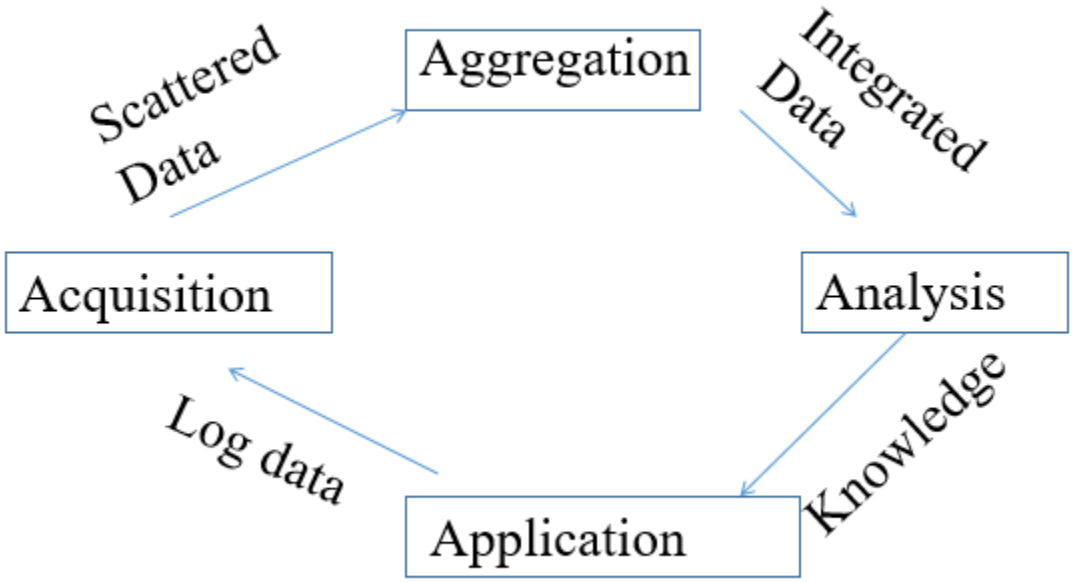
**X: Input Data and Y:Output Data**

| X | Y | **if X=10 then Y?** |
|---|---|---|
| 1 | 2 | |
| 2 | 4 | |
| 3 | 6 | |
| 4 | 9 | |
| 5 | 11 | |
| 6 | 11 | |
| 7 | 13 | |
| 8 | 17 | |
| 9 | 19 | |

# Example 1: Housing Data

```
Per_square_feet,price
150,6450
200,7450
250,8450
300,9450
350,11450
400,15450
600,18450
```

# Life Cycle of Data



# Life Cycle

**Acquisition**

- Acquiring Data from Various Data Sources.
- Raw data is collected from various logs, collected from various sources.

**Aggregation:**

- Raw data is gathered and expressed in a summary form for statistical analysis.
- average, minimum, maximum, sum, and count.

# Life Cycle

**Analysis:**

- Analyzing the aggregated data to gain insights about particular resources or resource groups.
- Deriving Knowledge from data

**Application:**

- Applying this knowledge in building applications

# Big Data Advantages

- Big data analysis derives innovative solutions. Big data analysis helps in understanding and targeting customers. It helps in optimizing business processes.
  ➡It helps in improving science and research.
  ➡It improves healthcare and public health with availability of record of patients.
  ➡It helps in financial tradings, sports, polling, security/law enforcement etc.

# Big Data Advantages

➡Any one can access vast information via surveys and deliver anaswer of any query.

➡Every second additions are made.

➡One platform carry unlimited information

# Big Data Disadvantages

- Following are the drawbacks or **disadvantages of Big Data**:

  ➡Traditional storage can cost lot of money to store big data.

  ➡Lots of big data is unstructured.

  ➡Big data analysis violates principles of privacy.

  ➡It can be used for manipulation of customer records.

  ➡It may increase social stratification.

# Big Data Disadvantages

➡Big data analysis is not useful in short run. It needs to be analyzed for longer duration to leverage its benefits.

➡Big data analysis results are misleading sometimes.

➡Speedy updates in big data can mismatch real figures.

# Types of Big Data

- **Social Networks (human-sourced information)**:
  - Social Networks: Facebook, Twitter, Tumblr etc.
  - Blogs and comments
  - Personal documents
  - Pictures: Instagram, Flickr, Picasa etc.
  - Videos: Youtube etc.
  - Internet searches
  - Mobile data content: text messages
  - User-generated maps
  - E-Mail

# Types of Big Data

- **Traditional Business systems (process-mediated data)**:

Data produced by Public Agencies

- Medical records

Data produced by businesses

- Commercial transactions
- Banking/stock records
- E-commerce
- Credit cards

# Types of Big Data

- **Internet of Things (machine-generated data)**:

Data from sensors

Fixed sensors

Home automation

Weather/pollution sensors

Traffic sensors/webcam

Scientific sensors

Security/surveillance videos/images

Mobile sensors (tracking)

Mobile phone location

Cars

Satellite images

# Types of Big Data

Data from computer systems

- Logs
- Web logs