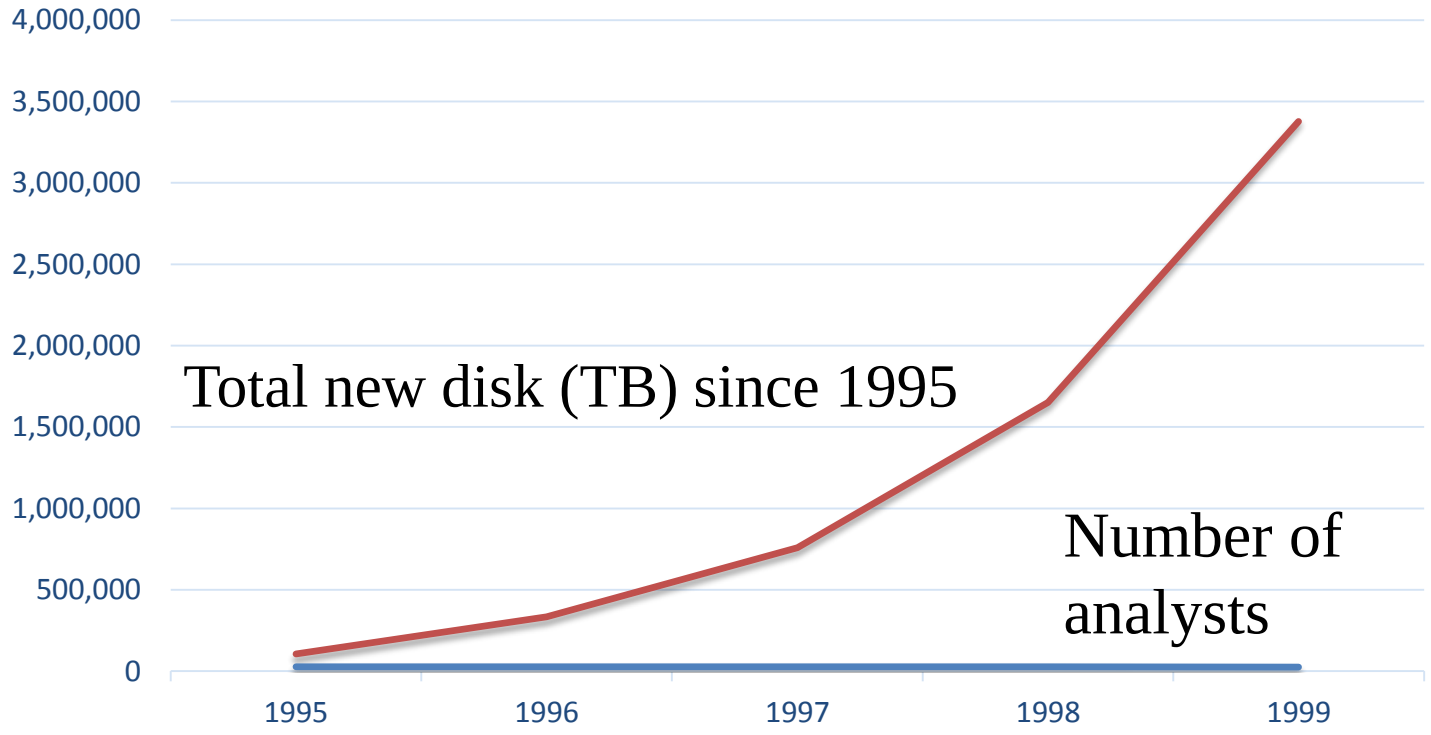
Data Mining

Farideh Bagherzadeh-Khiabani

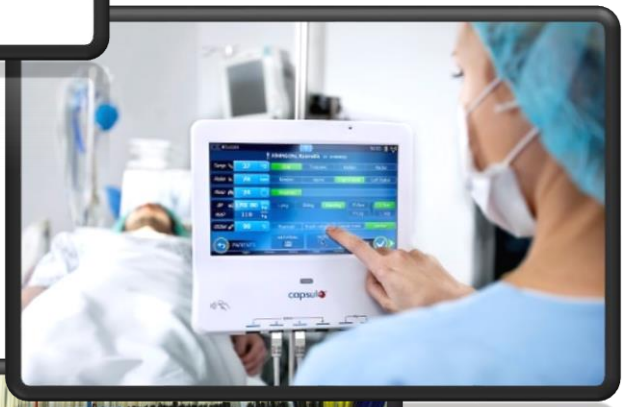
Prevention of Metabolic Disorders Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

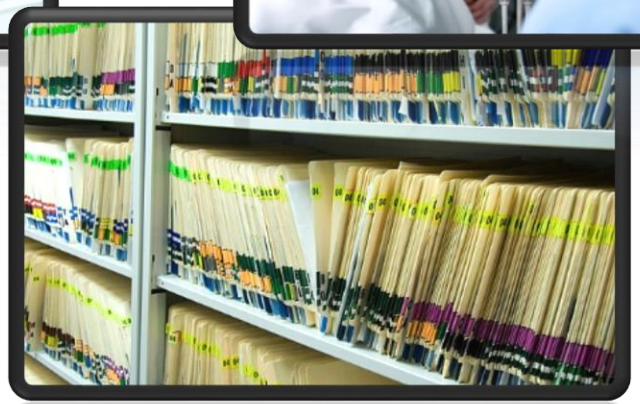
Why Data Mining?

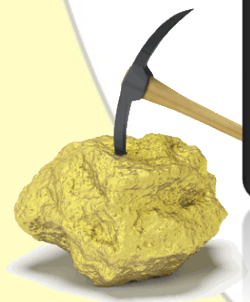


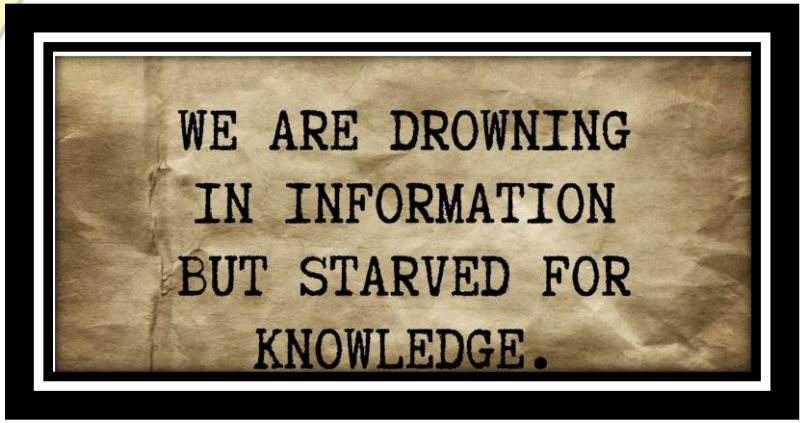
What Data Mining in Medicine?

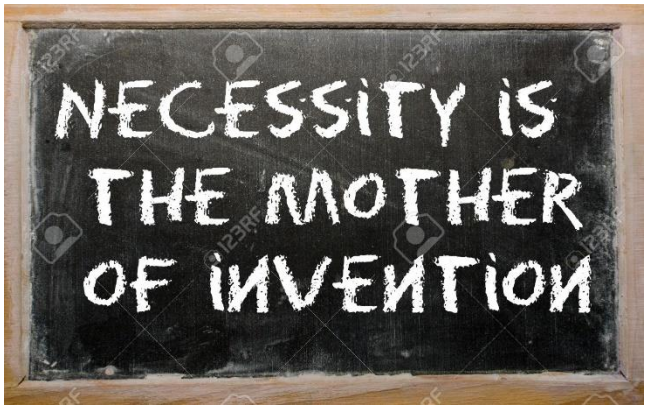
یک بیمارستان ممکن است
سالانه پنج ترابایت داده
تولید کند



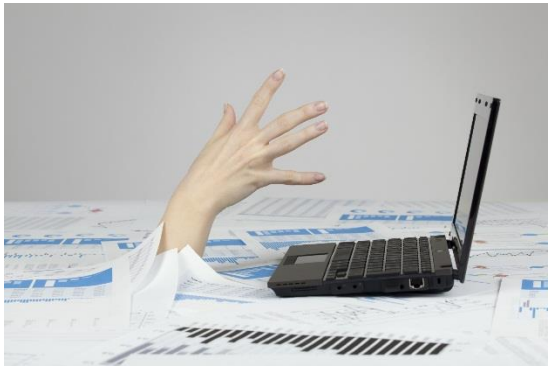




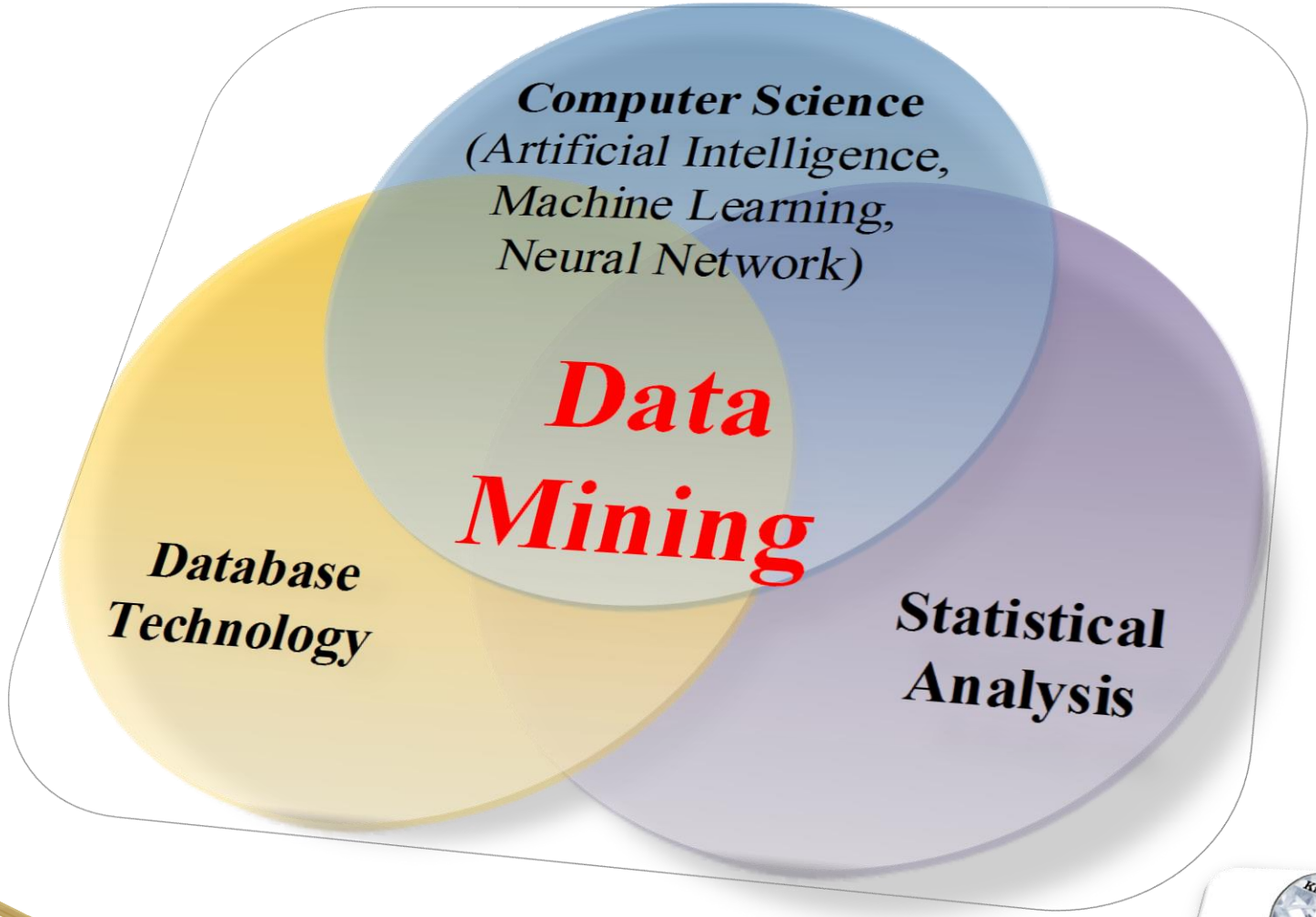

WE ARE DROWNING
IN INFORMATION
BUT STARVED FOR
KNOWLEDGE.

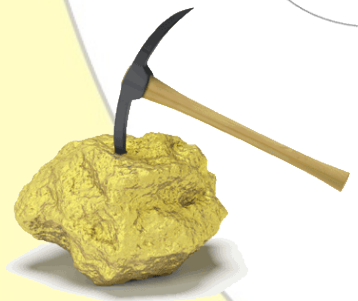

NECESSITY IS
THE MOTHER
OF INVENTION

HELP



Data mining

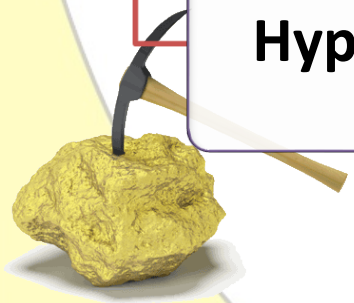

Computer Science
(Artificial Intelligence,
Machine Learning,
Neural Network)
Data
Mining
Database
Technology
Statistical
Analysis



Knowledge
Discovery

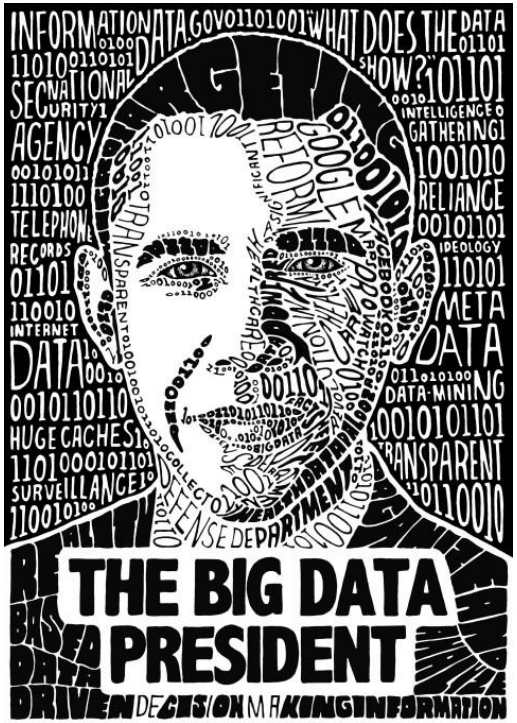
Data mining and Statistical Analysis

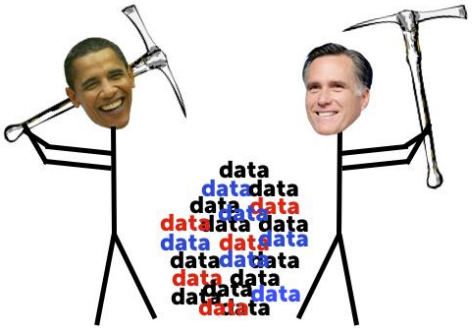
- High dimensional data
- Various types of data
- Hypothesis



"The secret of success is to know something that nobody else knows."

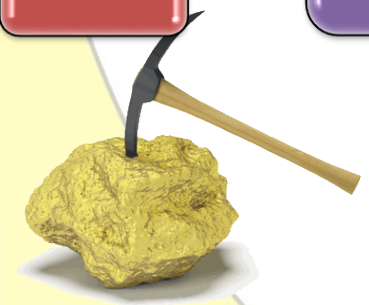
Aristotle Onassis





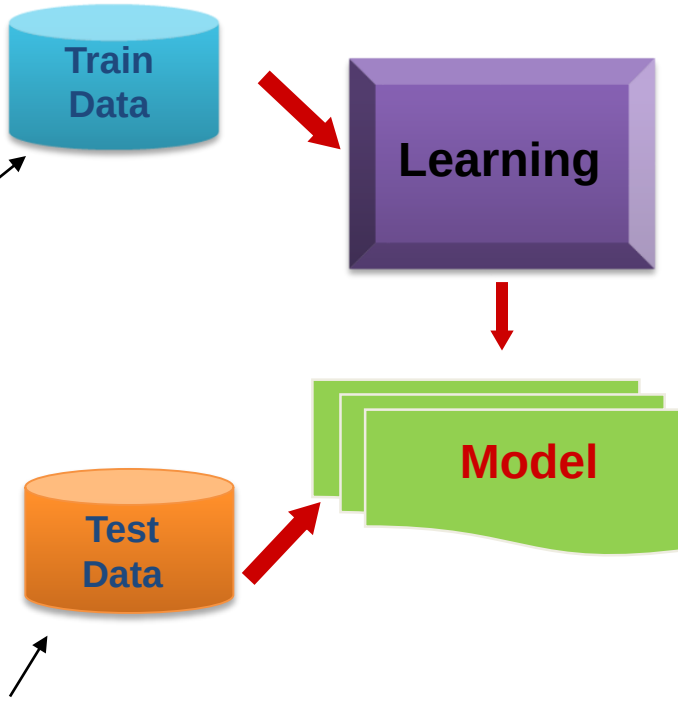
Data Mining Tasks

- Supervised (Predictive)
- Unsupervised (Descriptive)

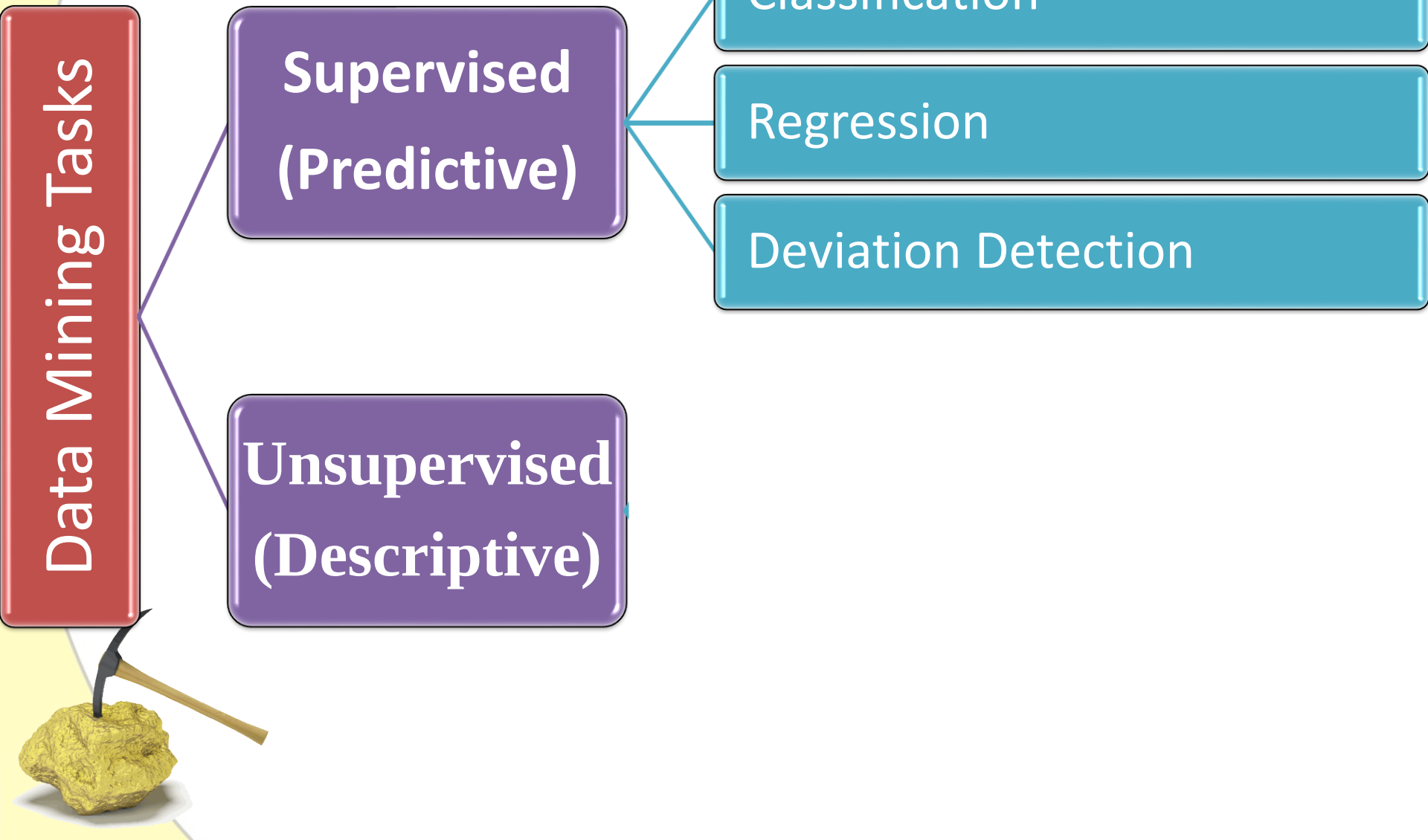


Supervised

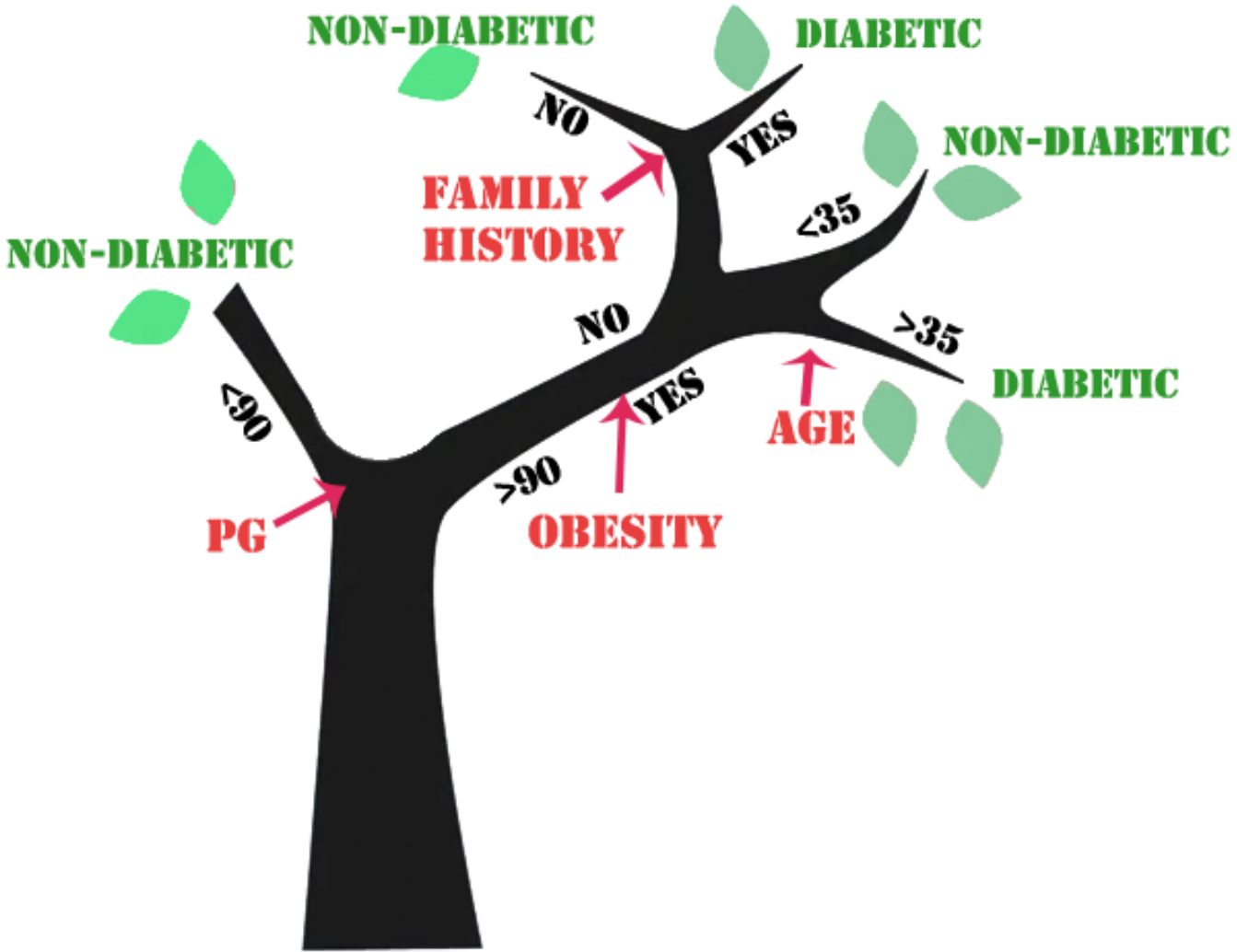
Code	Family history	Age	Wight	Diabetes Status
1	Yes	45	84	Yes
2	Yes	42	73	Yes
3	Yes	27	52	No
4	Yes	32	59	Yes
5	No	4	24	No
6	No	80	68	No
7	Yes	23	48	Yes
8	No	20	15	No



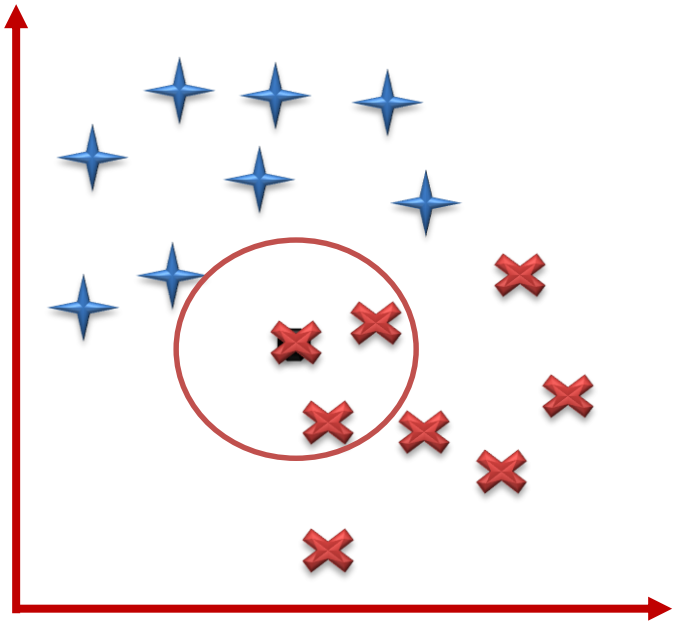
Code	Family history	Age	Wight	Diabetes Status
1	Yes	28	56	Yes
2	No	91	103	Yes
3	Yes	17	58	No


Data Mining Tasks
Supervised (Predictive)
Classification
Regression
Deviation Detection
Unsupervised (Descriptive)

Decision Tree



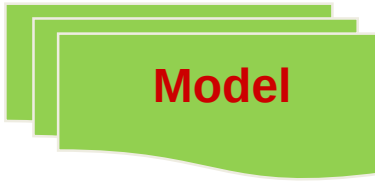
Nearest Neighborhood

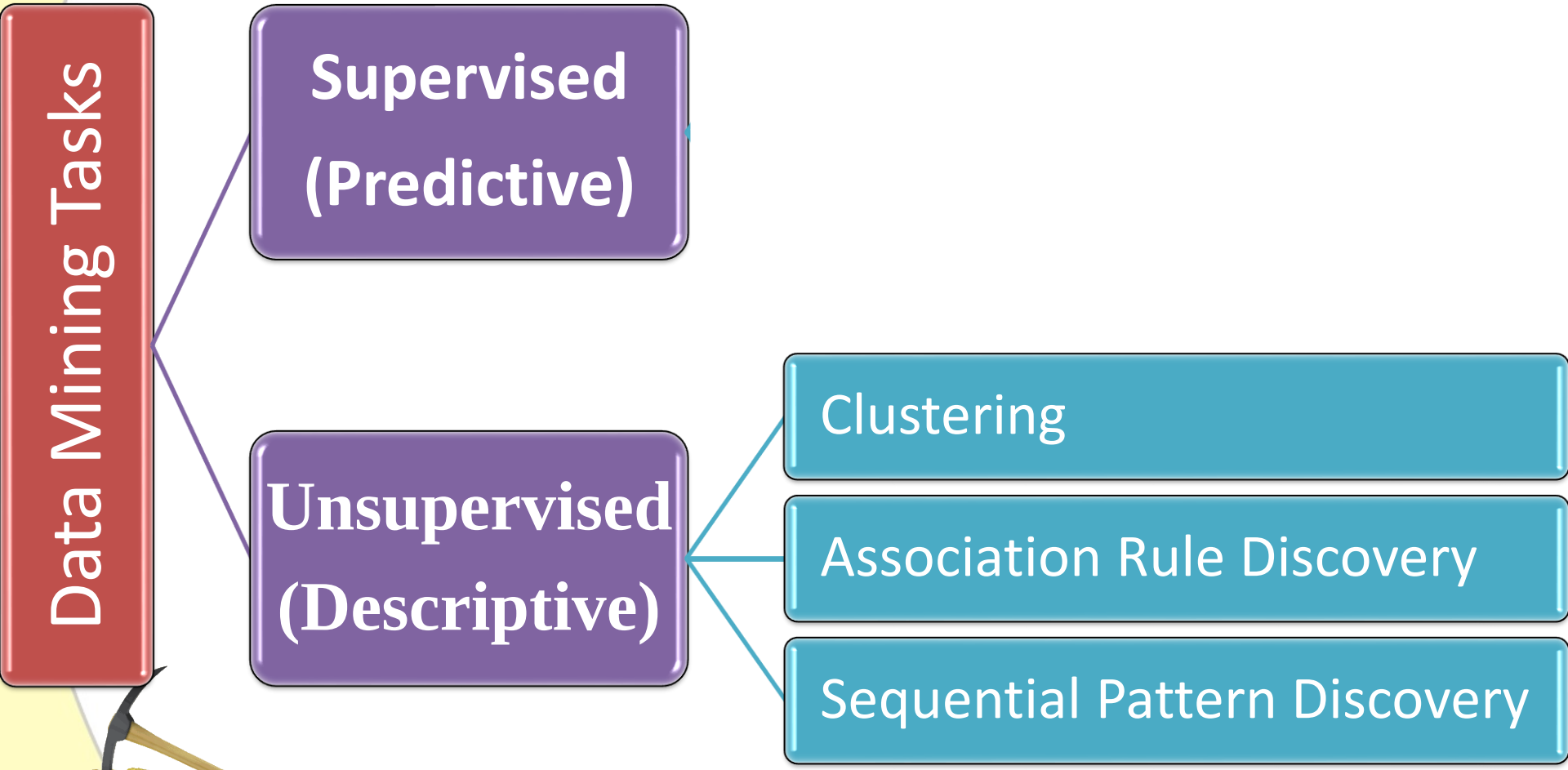


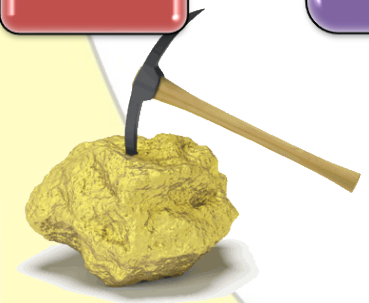
Un-Supervised

Code	Family history	Age	Wight	Diabetes Status
1	Yes	45	84	Yes
2	Yes	42	73	Yes
3	Yes	27	52	No
4	Yes	32	59	Yes
5	No	4	24	No
6	No	80	68	No
7	Yes	23	48	Yes
8	No	20	15	No

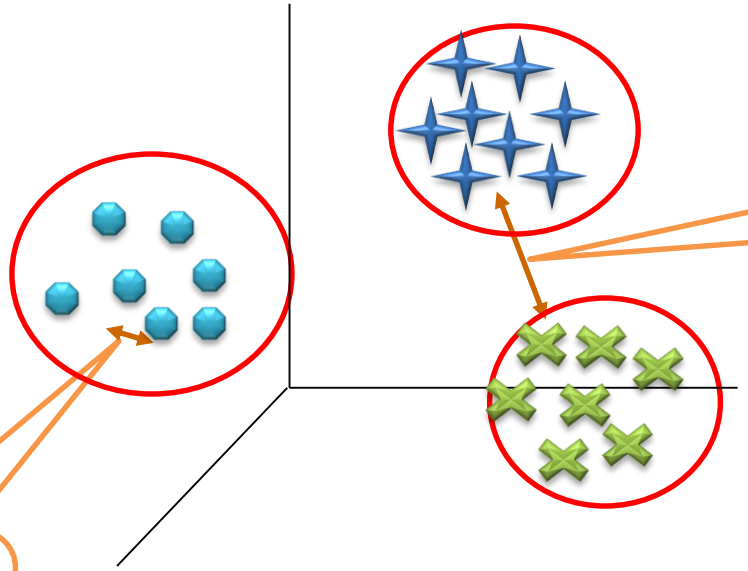
Code	Family history	Age	Wight	Diabetes Status
1	Yes	36	56	Yes
2	No	91	103	Yes
3	Yes	17	58	No




Data Mining Tasks
Supervised (Predictive)
Unsupervised (Descriptive)
Clustering
Association Rule Discovery
Sequential Pattern Discovery

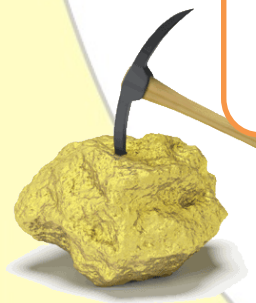


Clustering

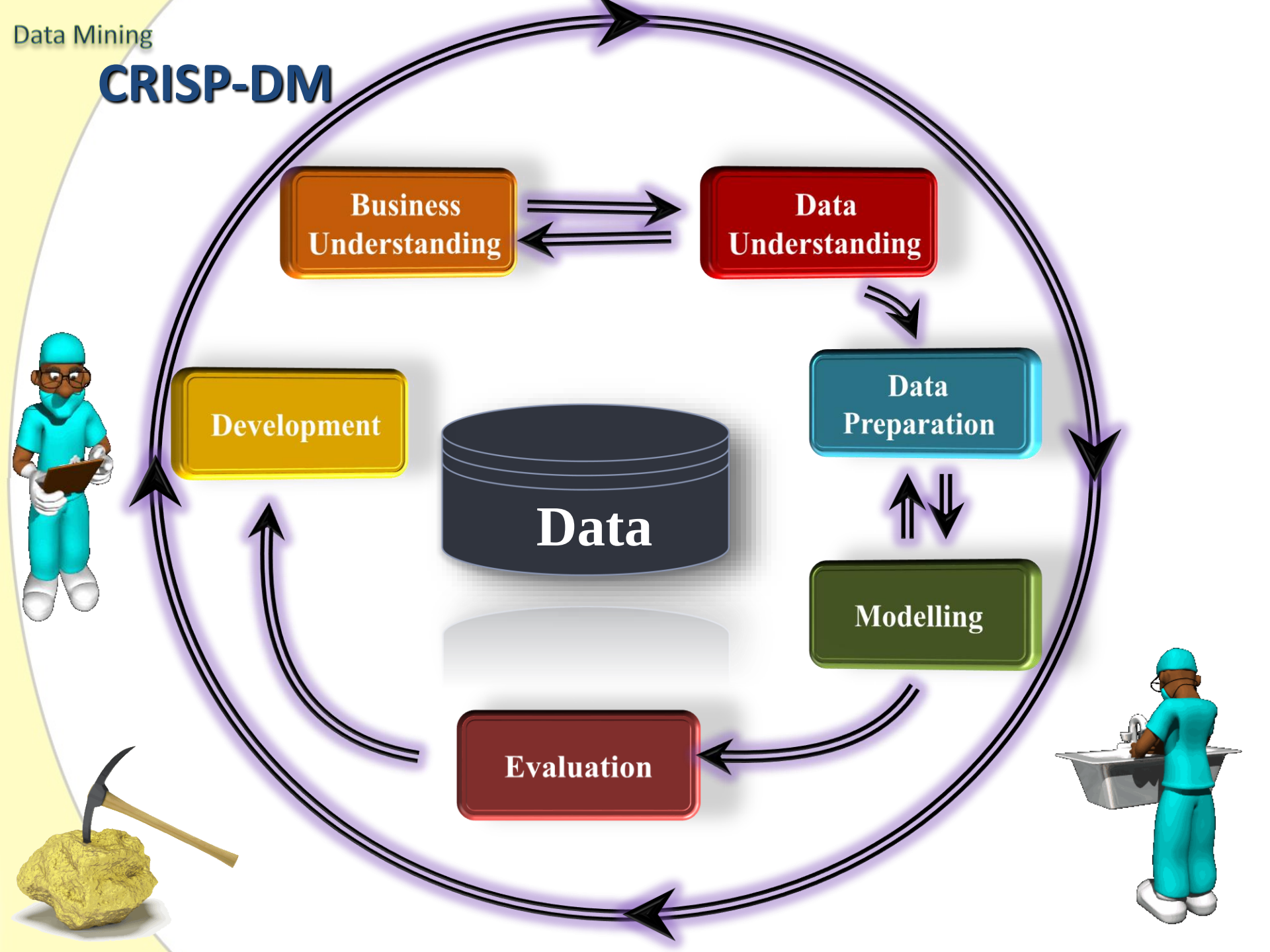
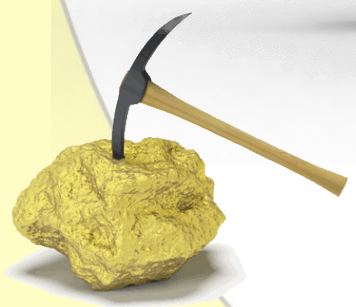


فاصله ی بین
خوشه ها
بیشینه شود

فاصله ی بین
اعضای خوشه
کمینه شود



CRISP-DM


Business Understanding
Data Understanding
Data Preparation
Modelling
Evaluation
Development
Data

Data Mining