Educational Challenges Related to Data Science Across Disciplines

Prof. Milena Stanković

Faculty of Electronic Engineering, University of Niš, Serbia

FUTURE EDUCATION CONFERENCE

Belgrade, November 11-13, 2019

Internet as a Global Resource

People, businesses, and devices have all become data factories that are pumping out incredible amounts of information to the web each day.

90% of the data on the internet has been created since 2016, according to an IBM Marketing Cloud study.

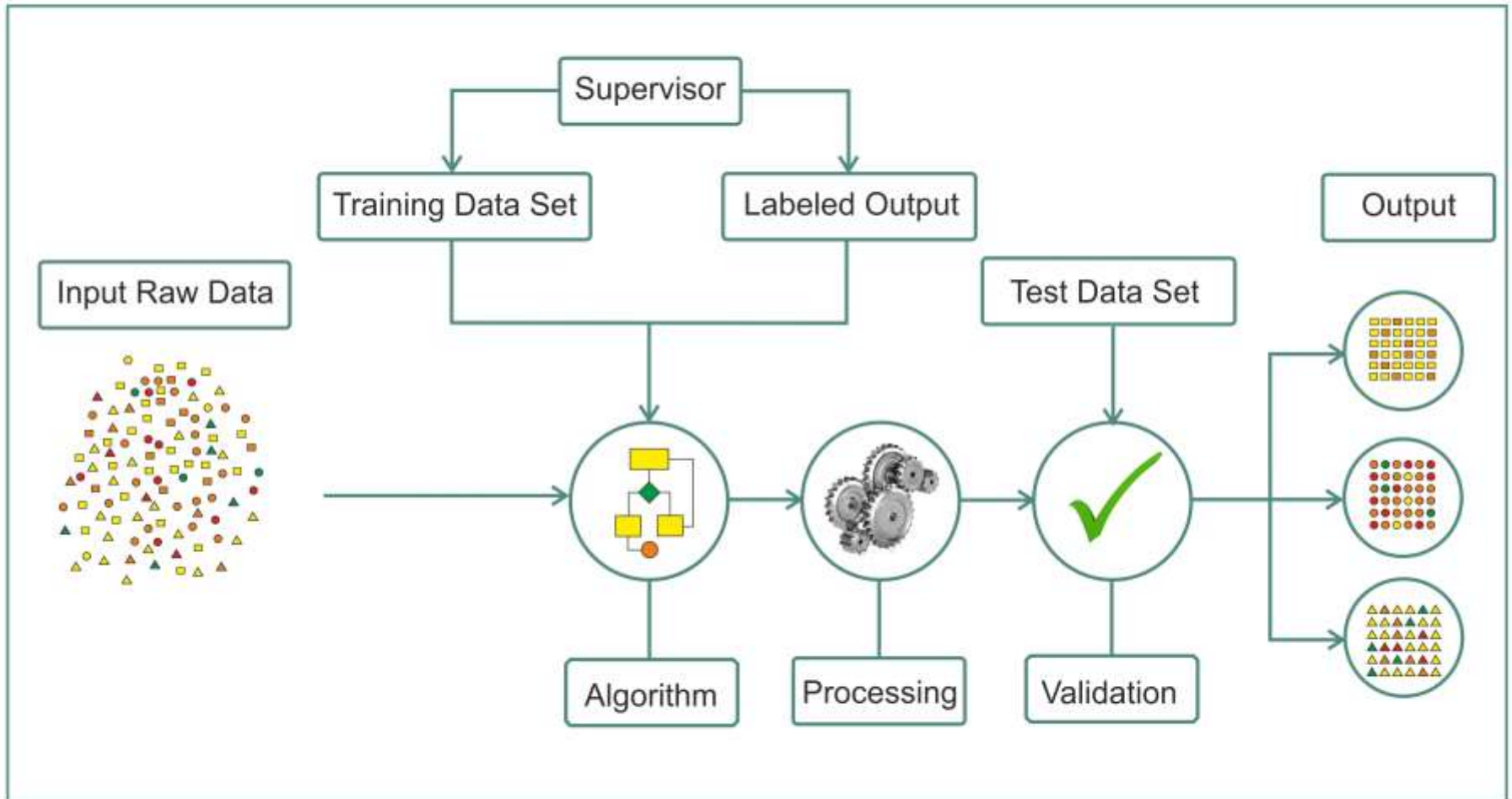**Zettabyte (1 000 000 000 000 000 000 000 Bytes)**
- According to IDC next year (2020) the total data on the internet will be 44 Zettabytes

Internet of Things

- The term generally refers to scenarios where network connectivity and computing capability extends to objects, sensors, and everyday items not normally considered computers.

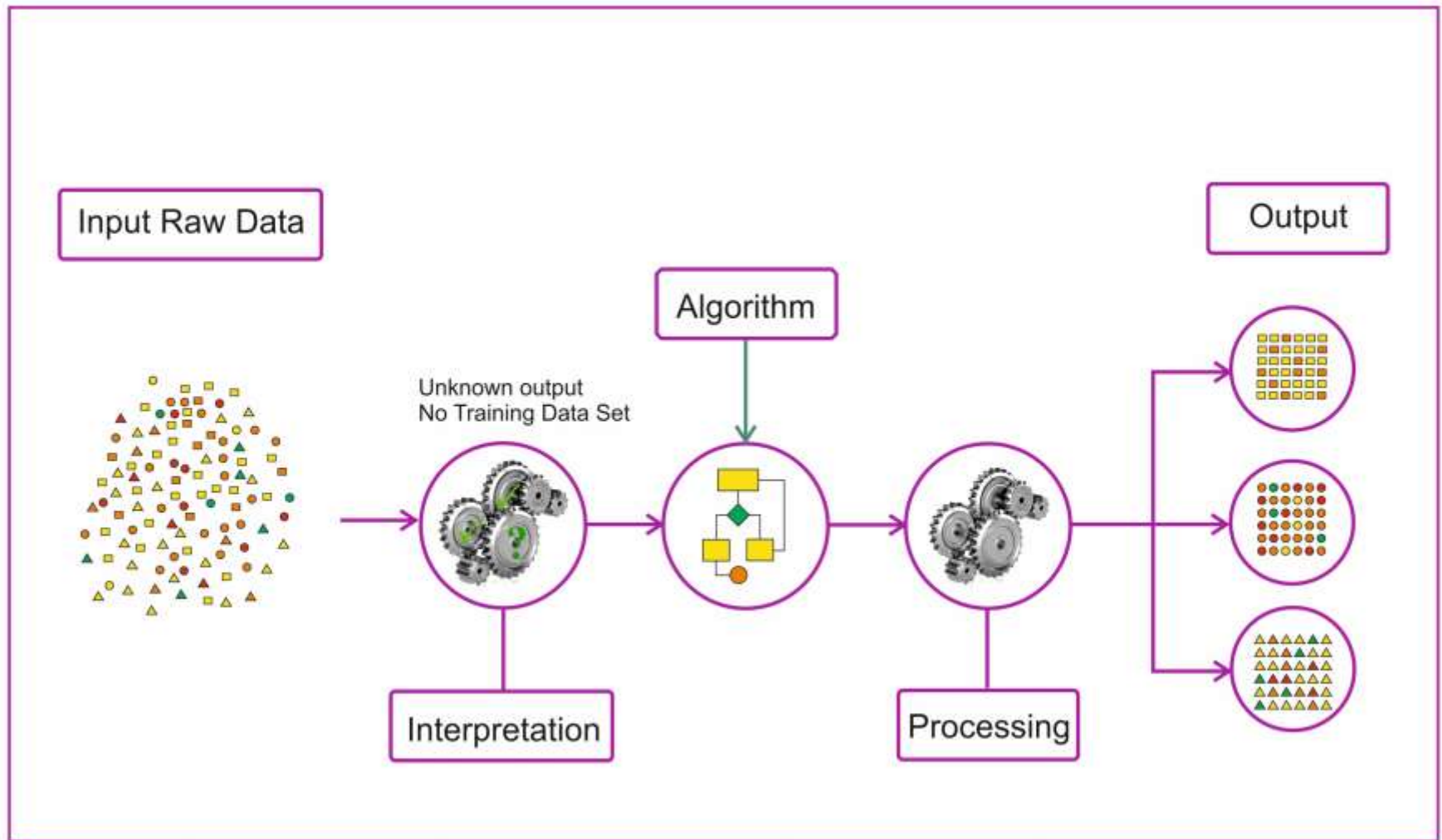- These devices generate, exchange, and consume data with minimal human intervention.

Unstructured Data

Traditional structured data, such as the transaction data in financial systems and other business applications, are adapted for processing and analyzing by machines.

Sets of unstructured data are not uniform. They are adapted to be used by humans, not by machines.
Examples:

- Unstructured text
- Images, audio, and video files
- Machine data, log files from websites, servers, networks, and applications – particularly mobile phones.

Supervised Machine Learning
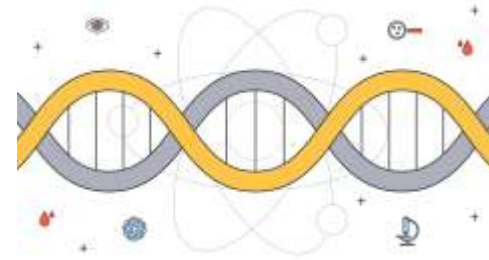## Classification

Unsupervised Machine Learning
Clustering

Data Science in Medicine and Healthcare

- Medical image analysis

- Genetics and genomics

- Predictive medicine: prognosis and diagnostic accuracy

- Virtual assistance for patients and customer support

- Creation of drugs

Data Science in Business

- Empowering management and officers to make better decisions
- Directing actions based on trends—which help to define goals and priorities
- Challenging the staff to adopt best practices and focus on issues that matter
- Identifying opportunities
- Testing of decisions
- **Fraud detection**: Banks use data science to determine if there is strange activity on some account.
- Identification and refining of target audiences
- Recommendation systems in e-commerce applications
- Recruiting the right talent for the organization
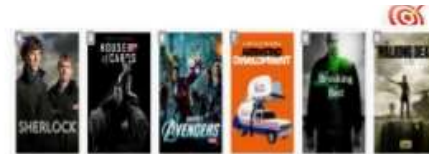
Recommendation Systems
Collaborative Filtering

User-based CF

Item-based CF

Target (or Active) user for whom the CF recommendation task is performed

sim(6,5) cannot be calculated

2			4	5	
5		4			1
		5		2	
	1		5		4
		4			2
4	5		1		

Alexandros Karatzoglou – September 06, 2013 – Recommender Systems

essir

NETFLIX

2			4	5	
5		4			1
		5		2	
	1		5		4
		4			2
4	5		1		

im(i,j): -1   -1   0.86   1   NA

Xavier Amatriain – July 2014 – Recommender Systems

Data Science in Industrial Processes

Quality control

Anomaly detection (outliers)

Personalized products and services

Power consumption prediction

Cost reduction

Service oriented production

INDUSTRY 4.0

INDUSTRY 4.0

Today
Cyber physical sys.
IoT, networks

INDUSTRY 3.0

1969 Automatization
Computers and electronic

INDUSTRY 2.0

1870 Electrical energy
Mass production

INDUSTRY 1.0

1784 Mechanization
Steam power

Data Engineer

What he knows
- Analytics
- Predictive modeling
- Statistical analysis and modeling
- Data mining
- Machine learning
- Sentiment analysis
- What-if analysis
- Visualization and presentation

After oil and gas geologists, data mining is the second highest paid job.

Duty
- Cleanse existing raw data and build models to predict future data
- Look at data from multiple angles and give meaning to it
- Identify correct bussines problems and offer solutions by best aplying the data

Impact to Organizations
- Develop strategies
- Improve operational efficiency
- Reduce cost
- Offer personalized product and services
- Mitigate risk

MATHEMATIC AND STATISTIC

- Statistical modeling
- Machine Learning
- Classification
- Clustering
- Time series prediction
- Dimensionality reduction
- Theory of optimization
- Experiment design

DOMAIN KNOWLEDGE AND SOFT SKILLS

- Critical thinking
- Proactive problem solving
- Intellectual curiosity
- Business sense
- Effective communication

SOFTWARE ENGINEERING

- Computer Science fundamentals
- Programing Language (Payton)
- R programming
- Relational and non-relational Databases
- Big data tools: Hadoop …
- GP GPU Programming
- Claud Computing

VISUALIZATION AND PREZETATION SKILLS

- Information visualization
- Visual analytics
- Graphic
- Design aesthetics
- Visualization solutions and tools

Education

- Open and flexible education system

- Multidisciplinary: Mathematic, Programming, Soft skills, Arts

- Development of capacities for independent and creative thinking, and problem solving.

- Collaborative, project oriented, education which prepares for critical thinking and problem solving

- **Prepare young generation for exponential growing of technologes.**