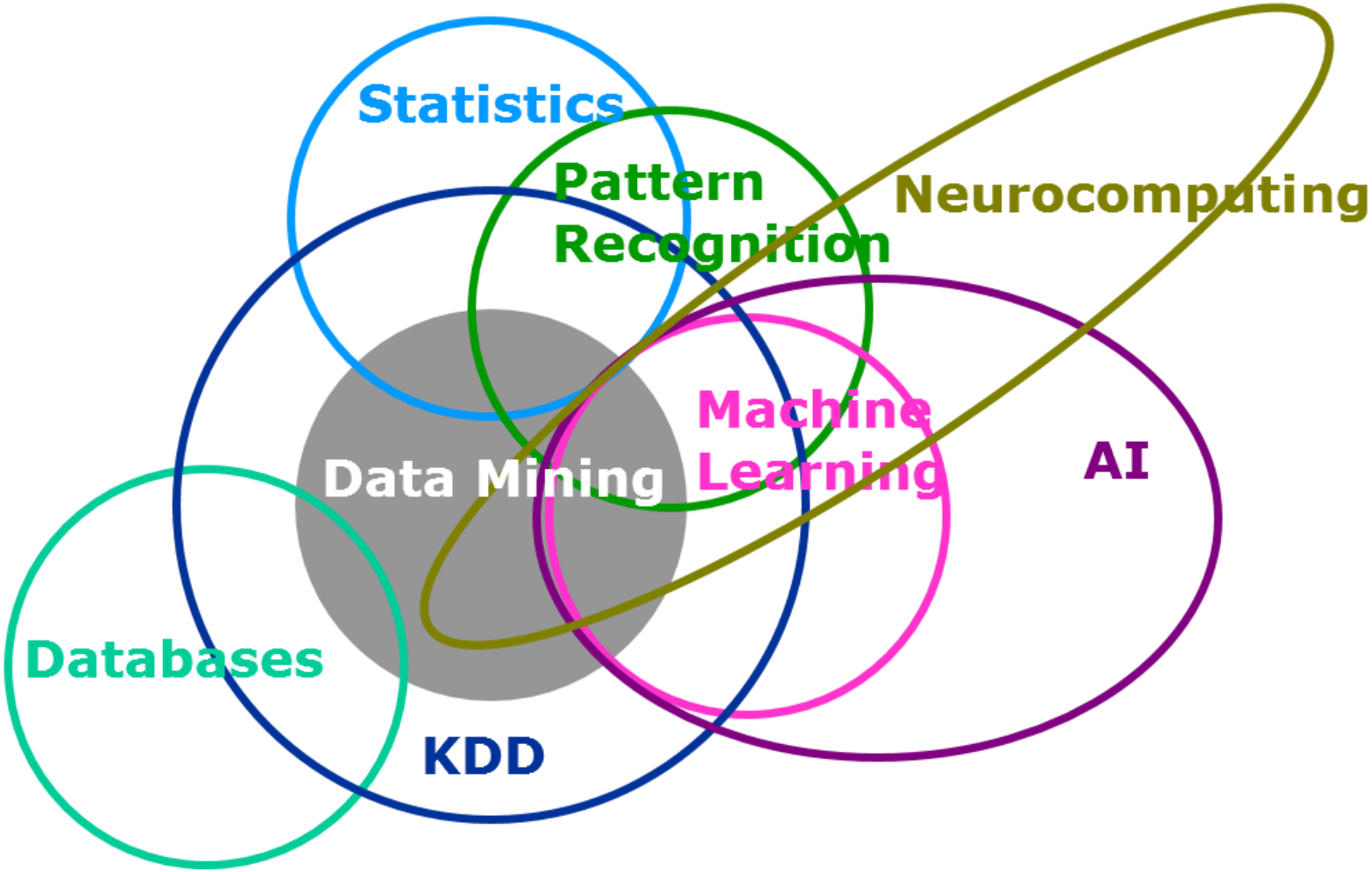
Puzzle words aka Terminology

Venn Diagram of Data mining



Data Mining the Venn Diagram by SaS

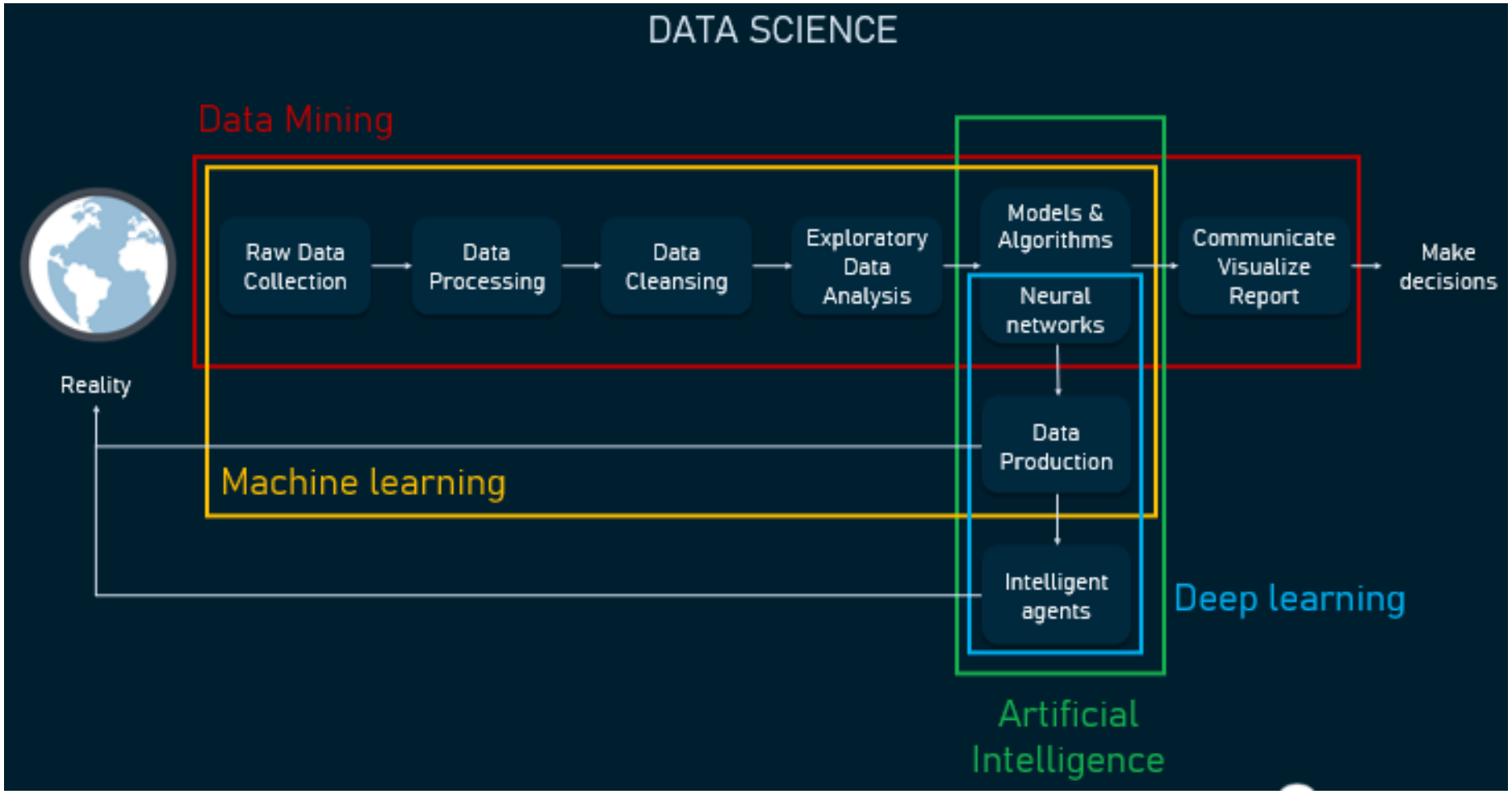
AI vs ML vs DL vs ML

 Artificial Intelligence	 Machine Learning	 Deep Learning
Artificial intelligence originated around 1950s.	Machine learning originated around 1960s.	Deep learning originated around 1970s.
AI represents simulated intelligence in machines.	Machine Learning is the practice of getting machines to make decisions without being programmed.	Deep Learning is the process of using Artificial Neural Networks to solve complex problems.
AI is a subset of Data Science.	Machine learning is a subset of AI & Data Science	Deep learning is a subset of Machine learning, AI & Data Science.
Aim is to build machines which are capable of thinking like humans.	Aim is to make machines learn through data so that they can solve problems.	Aim is to build neural networks that automatically discover patterns for feature detection.

AI vs ML vs DL vs ML

- **Data mining:**
  - Explore that data to automatically extract the rules and relations from large quantities of data
  - Usually, to help with human decision making.
- **Machine learning:**
  - Using computer to automatically detect patterns in data and use these to make predictions or decisions
  - Most useful when:
    - We want to automate something a human can do.
    - We want to do things a human can't do (look at 1 TB of data).
- **Data mining and machine learning are very similar:**
  - Data mining often viewed as closer to databases.
  - Machine learning often viewed as closer AI

AI vs ML vs DL vs ML



https://content.altexsoft.com/media/2021/01/Data-science-and-AI.png

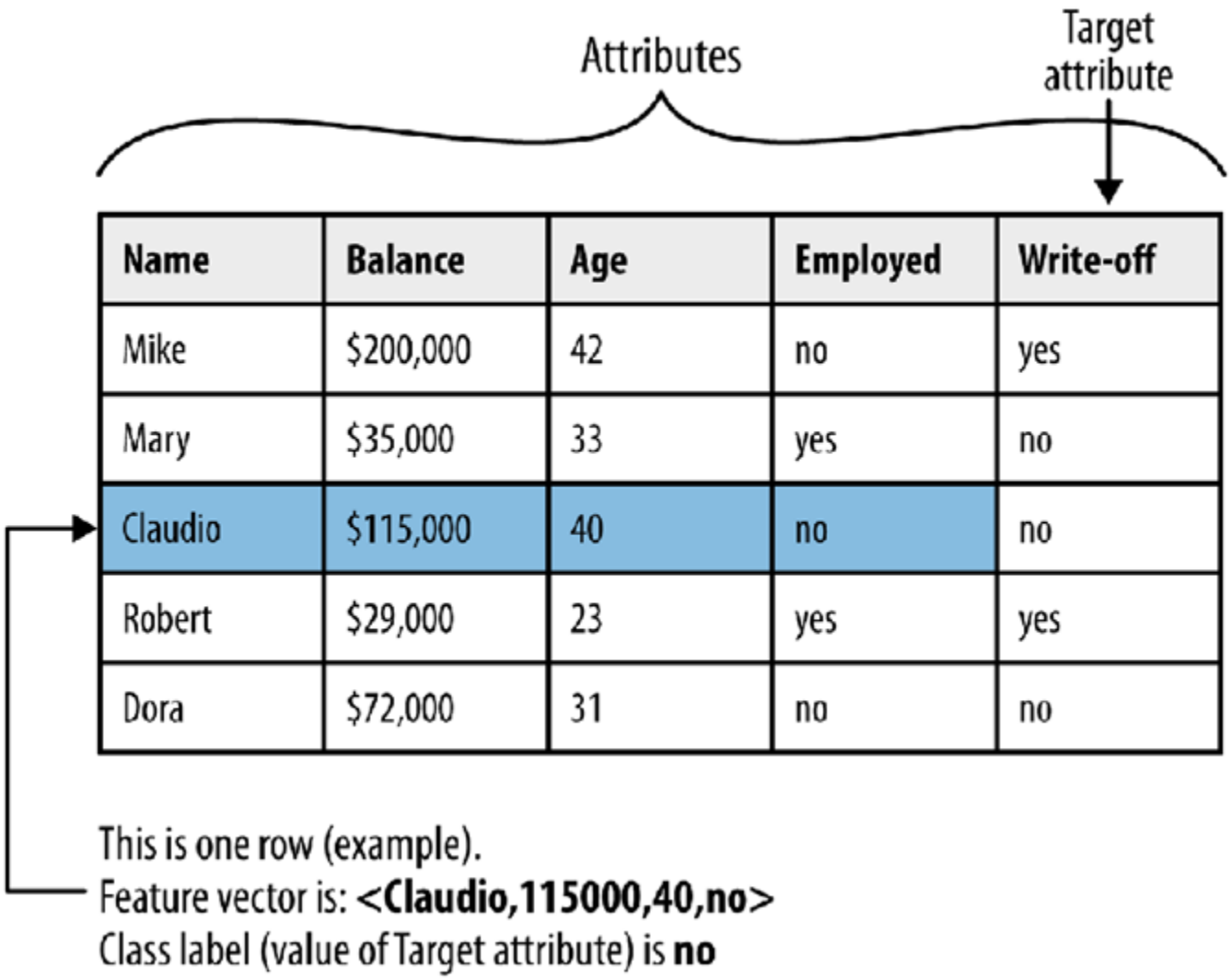
ML vs Stat

- In his course on statistics, Rob Tibshirani, a statistician who also has a foot in machine learning, provides a glossary that maps terms in statistics to terms in machine learning, reproduced below.

Glossary

Machine learning	Statistics
network, graphs	model
weights	parameters
learning	fitting
generalization	test set performance
supervised learning	regression/classification
unsupervised learning	density estimation, clustering
large grant = $1,000,000	large grant= $50,000
nice place to have a meeting: Snowbird, Utah, French Alps	nice place to have a meeting: Las Vegas in August

ML Terminology



ML Terminology

- Attribute (field, variable, feature): A quantity describing an instance.
- An attribute has a domain defined by the attribute type (possible values) :
  - Categorical (qualitative): A finite number of discrete values.
    - The type nominal denotes that there is no ordering between the values, such as last names and colors.
    - The type ordinal denotes that there is an ordering, such as in an attribute taking on the values low, medium, or high.
  - Continuous (quantitative): Commonly, subset of real numbers, where there is a measurable difference between the possible values. Integers are usually treated as continuous in practical problems.

ML Terminology

- A feature is the specification of an attribute and its value.
  - ``Color is blue'' is a feature of an example.
  - Many transformations to the attribute set leave the feature set unchanged (for example, regrouping attribute values or transforming multi-valued attributes to binary attributes).
  - Some authors use feature as a synonym for attribute.

- Dataset: A schema and a set of instances matching the schema.

- Instance (example, case, record, sample): A single object of the world from which a model will be learned, or on which a model will be used (e.g., for prediction).
  - Instances are described by feature vectors;
  - Some work uses more complex representations (e.g., containing relations between instances or between parts of instances).