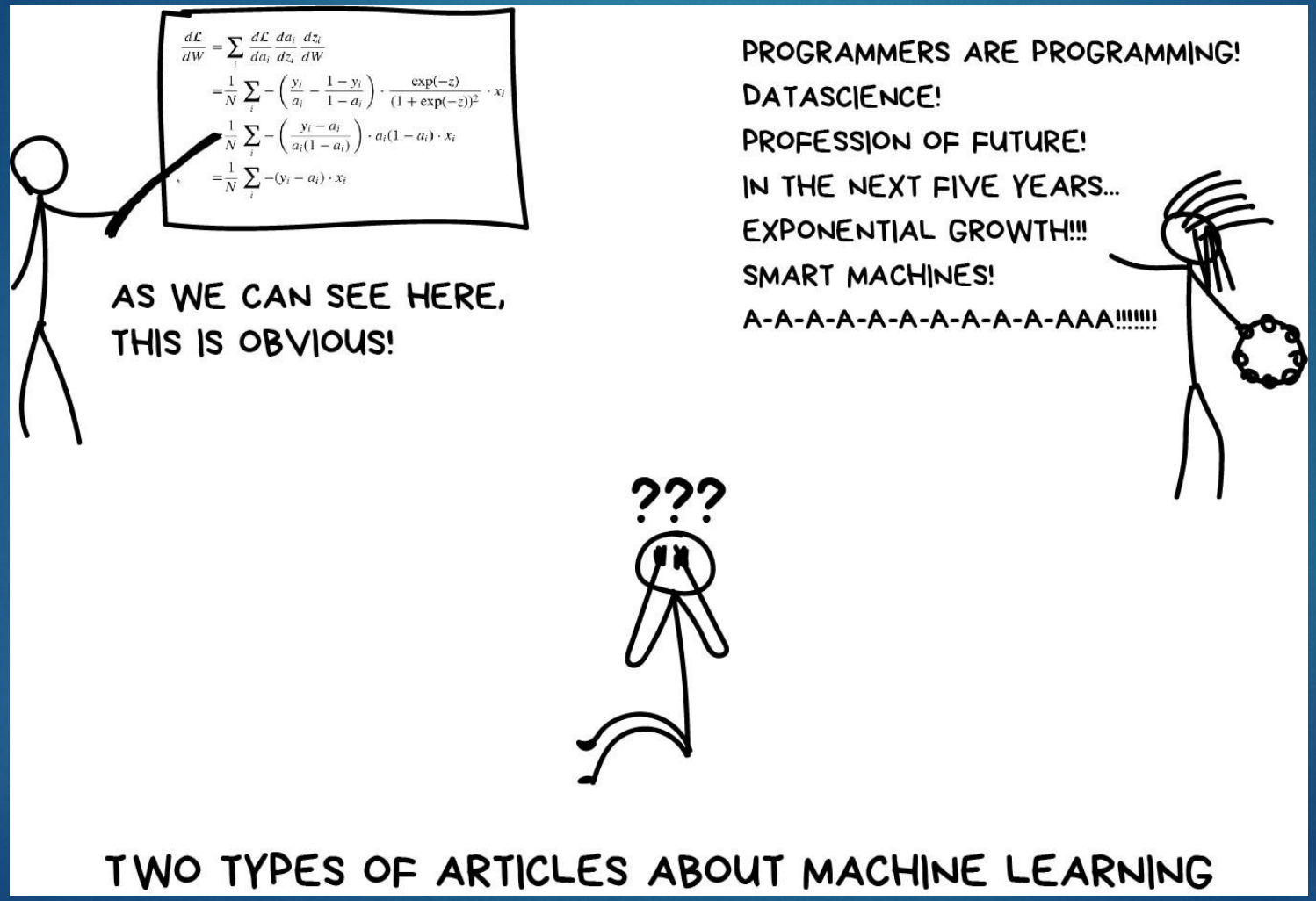
Machine Learning

What it is and Why use it.

Machine Learning

- If you ever tried to read articles about machine learning on the Internet, most likely you stumbled upon two types of them:
  - thick academic trilogies filled with theorems or
  - fishy fairytales about *artificial intelligence*, *data-science magic*, and *jobs of the future*.

Machine Learning



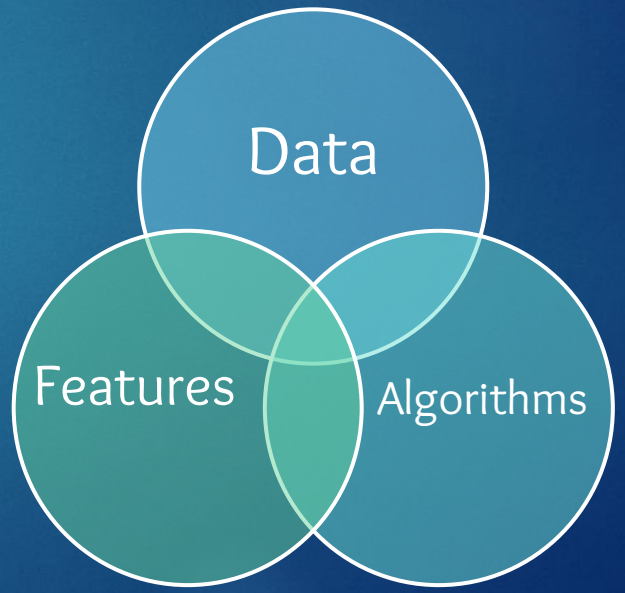
Machine Learning

- Machine Learning is the science (and art) of programming computers so they can *learn from data*.

- Machine learning is a powerful artificial intelligence tool that enables us to crunch petabytes of data and make sense of a complicated world.

Three Components of Machine Learning

- The only goal of machine learning is to predict results based on incoming data.
- That's it. All ML tasks can be represented this way, or it's not an ML problem from the beginning.
- The greater variety in the samples you have, the easier it is to find relevant patterns and predict the result.
- Therefore, we need three components to teach the machine:
  - Data
  - Features
  - Algorithms

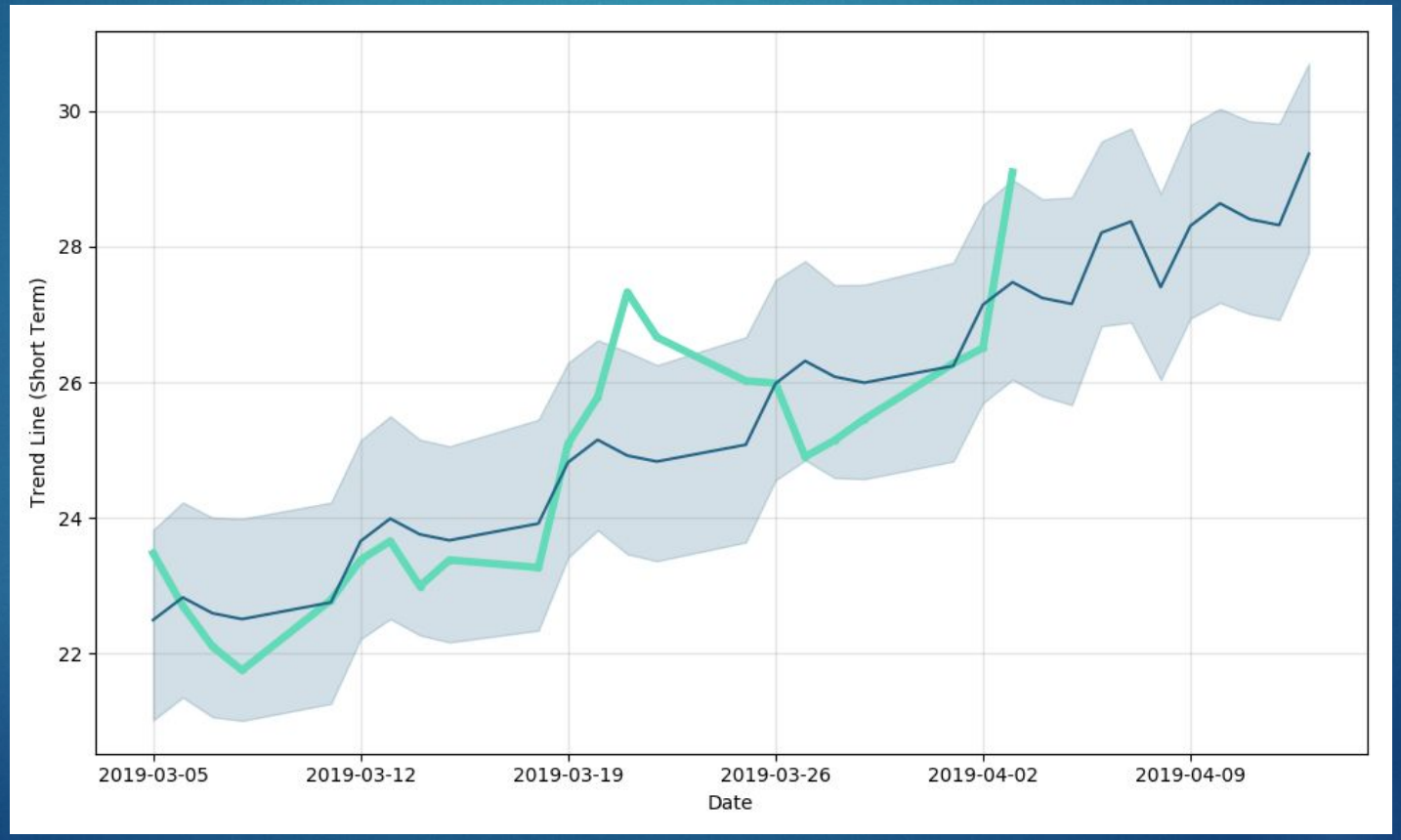


Data

- Want to detect spam? Get samples of spam messages.

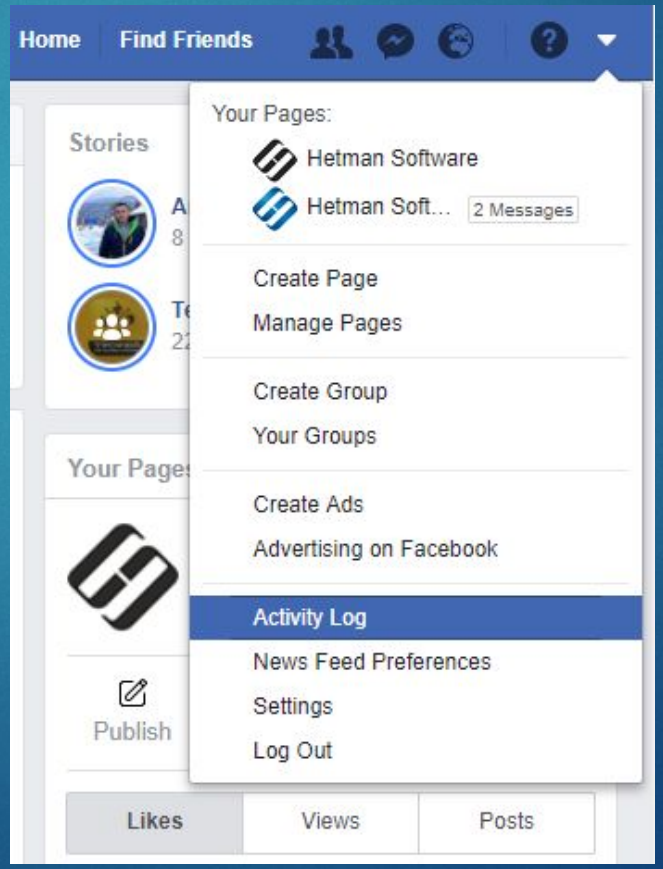
Data

- Want to forecast stocks? Find the price history.



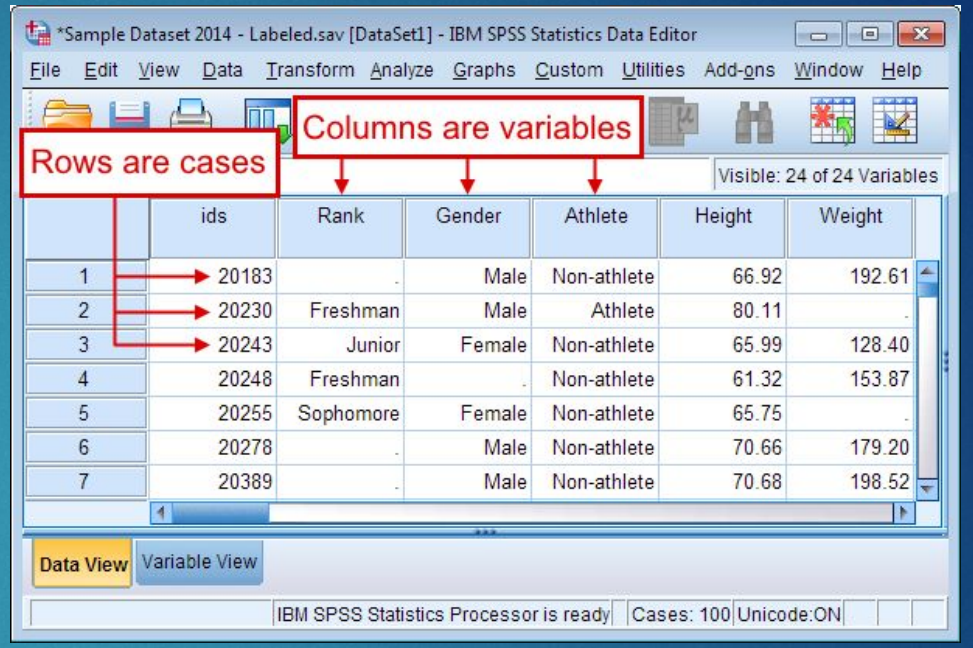
Data

- Want to find out user preferences? Parse their activities on Facebook.



Features

- Also known as parameters or variables.
- Those could be
  - car mileage,
  - user's gender,
  - stock price,
  - word frequency in the text.
- In other words, these are the factors for a machine to look at.
- When data is stored in tables, it's simple — features are column names.



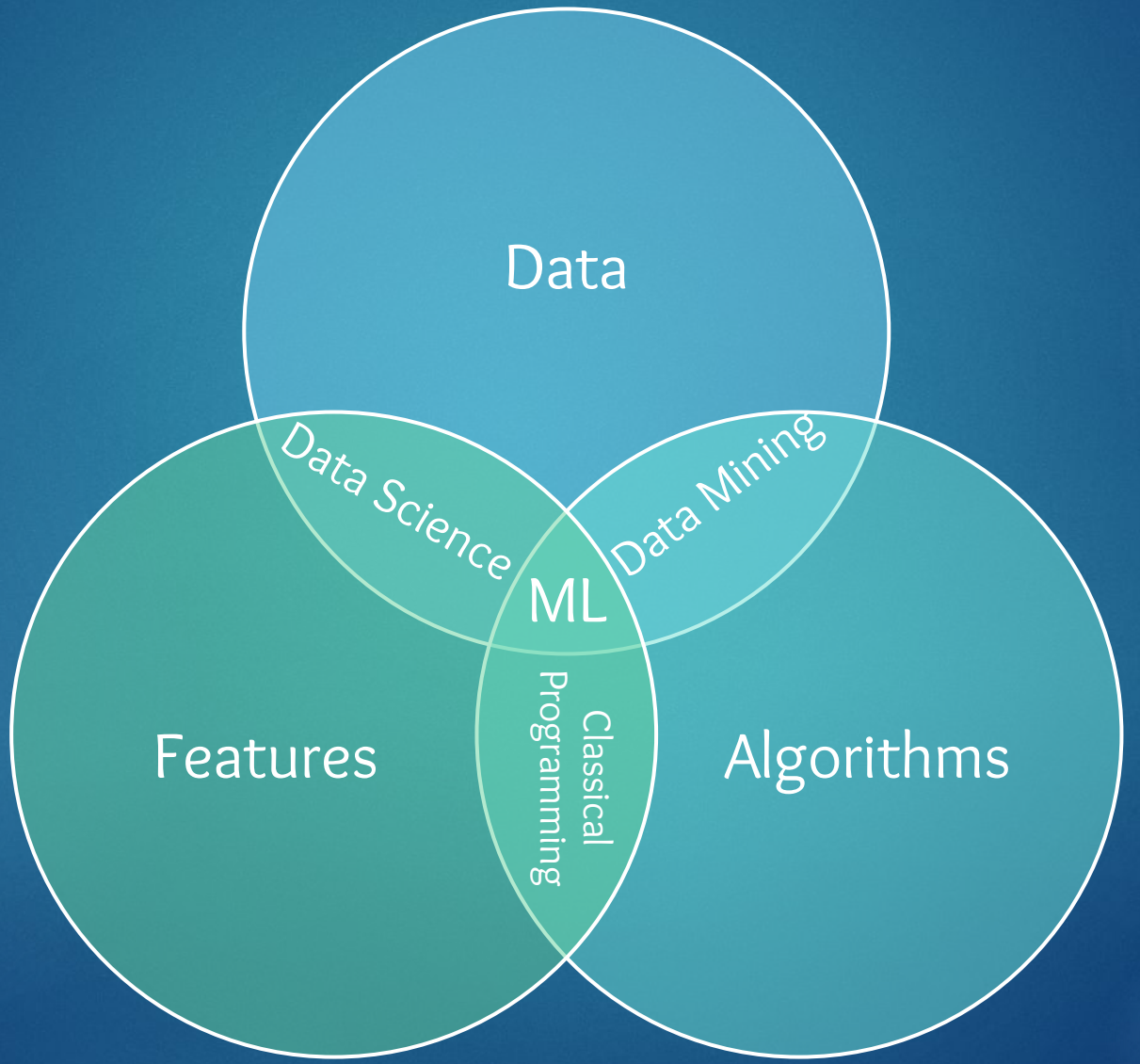
Features

- But what are they if you have 100 Gb of cat pics?
- We cannot consider each pixel as a feature.
- That's why selecting the right features usually takes way longer than all the other ML parts.
- That's also the main source of errors.
- Humans are always subjective.
- We choose only features we like or find "more important".
- Please, avoid being human.

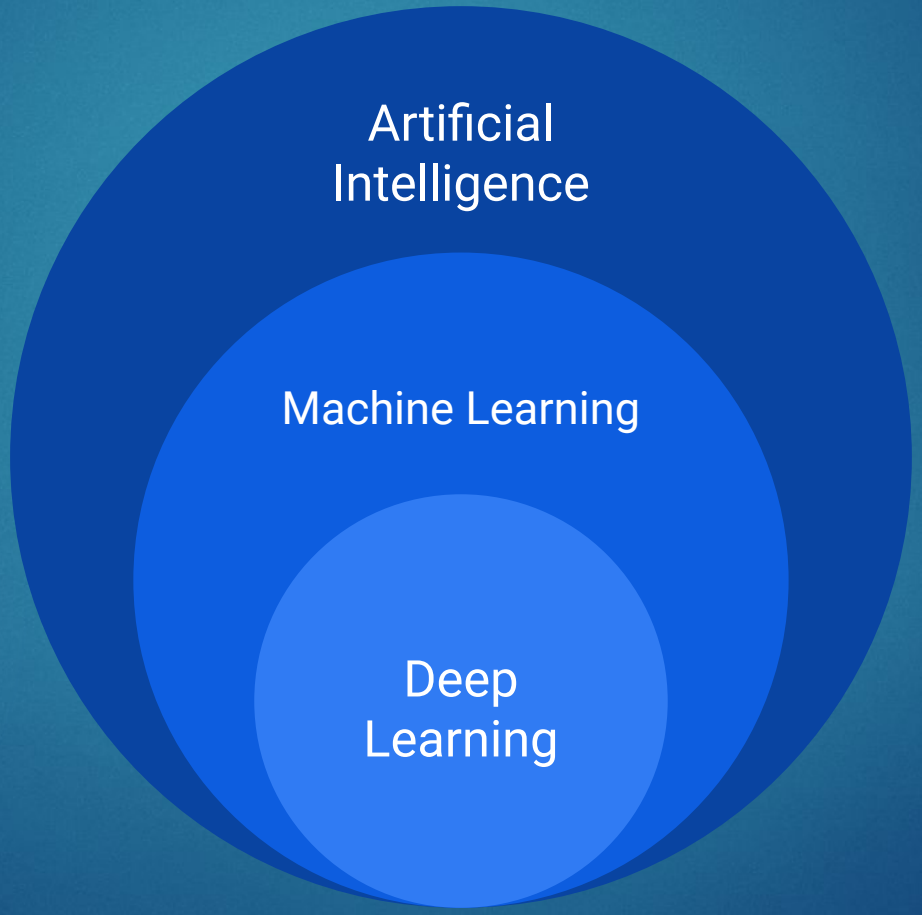
Algorithms

- Most obvious part. Any problem can be solved differently.
- The method you choose affects the precision, performance, and size of the final model.
- There is one important nuance though: if the data is crappy, even the best algorithm won't help.
- Sometimes it's referred as "garbage in – garbage out".
- So don't pay too much attention to the percentage of accuracy, try to acquire more data first.

Three Components of Machine Learning



Learning vs Machine Learning



Learning vs Machine Learning

Artificial Intelligence

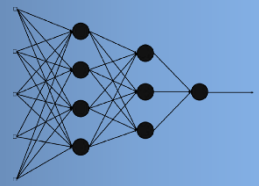
Engineering of making Intelligent Machines

Machine Learning

Ability to Learn without Programming

Deep Learning

Makes computation of multi-layer neural networks feasible



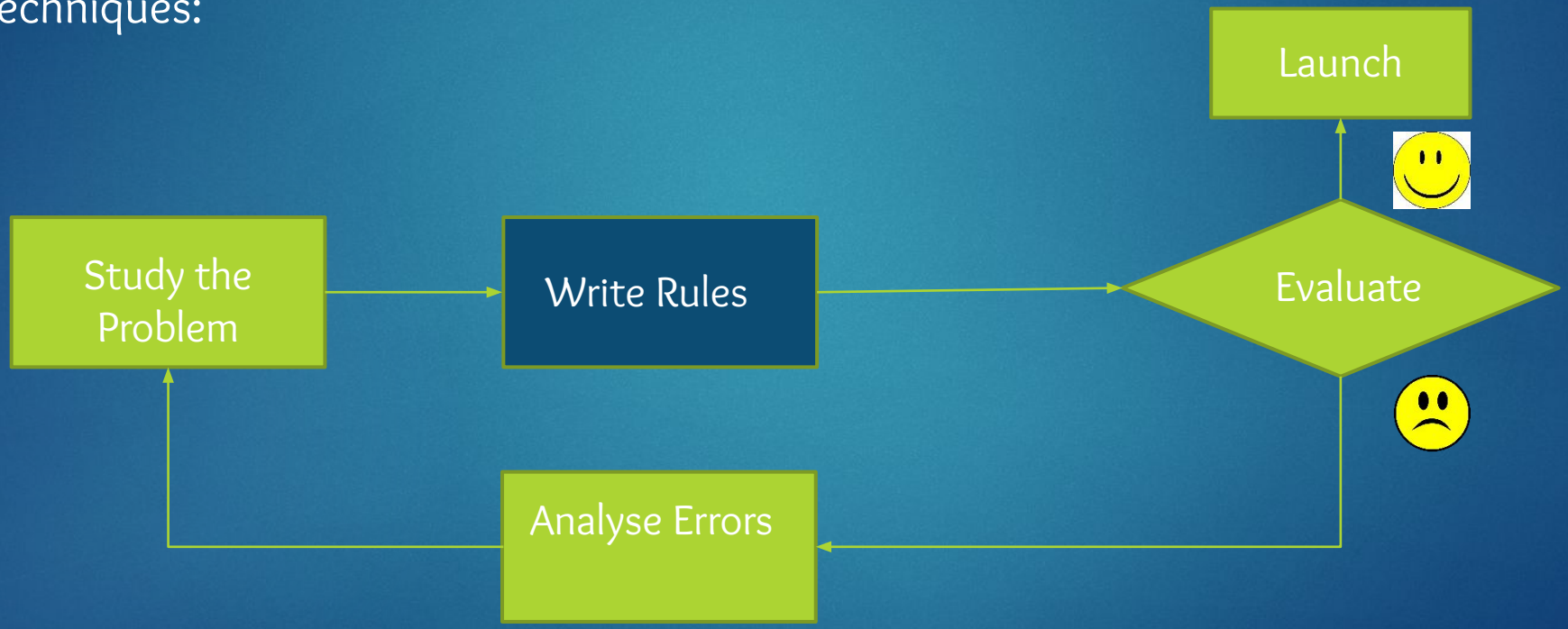
1950's 1960's 1970's 1980's 1990's 2000's 2010's

Learning vs Machine Learning

Machine can	Machine cannot
Forecast	Create something new
Memorize	Get smart really fast
Reproduce	Go beyond their task
Choose best item	Kill all humans

Why Use Machine Learning?

- Consider how you would write a spam filter using traditional programming techniques:

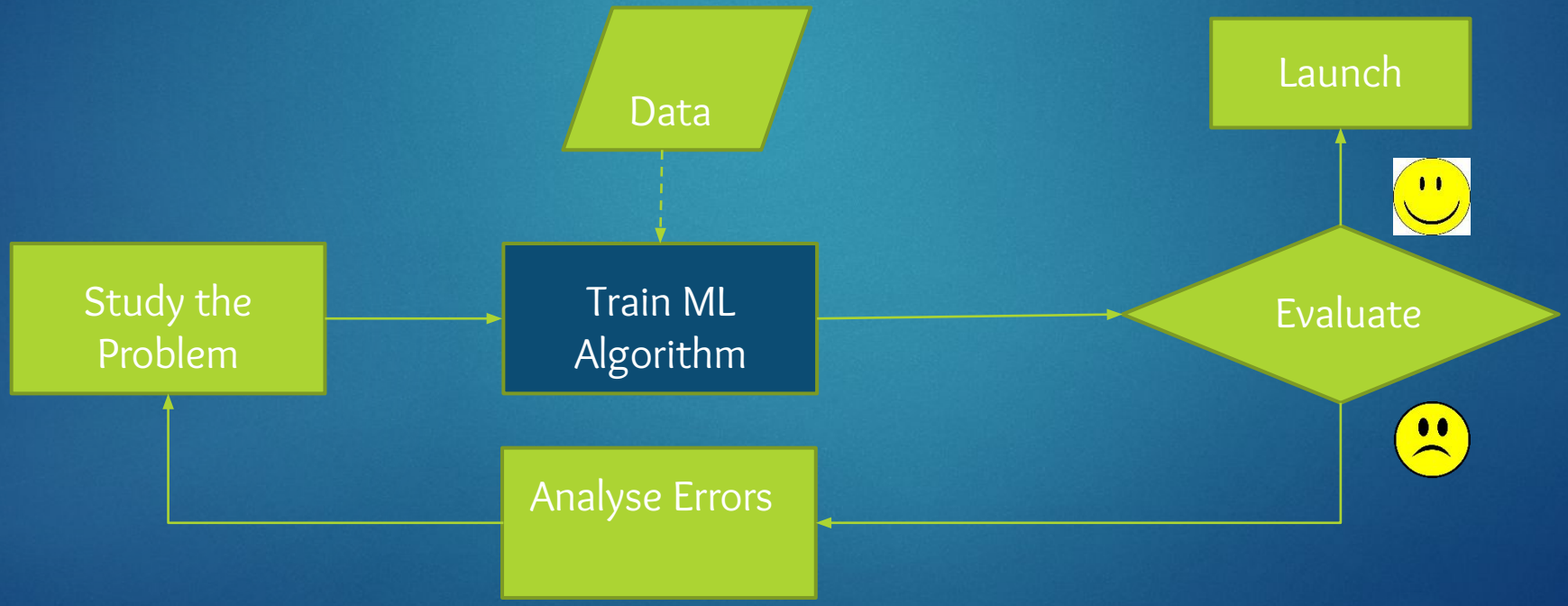


Why Use Machine Learning?

1. First you would look at what spam typically looks like.
   - You might notice that some words or phrases (such as “4U,” “credit card,” “free,” and “amazing”) tend to come up a lot in the subject.
   - Perhaps you would also notice a few other patterns in the sender’s name, the email’s body, and so on.
2. You would write a detection algorithm for each of the patterns that you noticed, and your program would flag emails as spam if a number of these patterns are detected.
3. You would test your program, and repeat steps 1 and 2 until it is good enough.

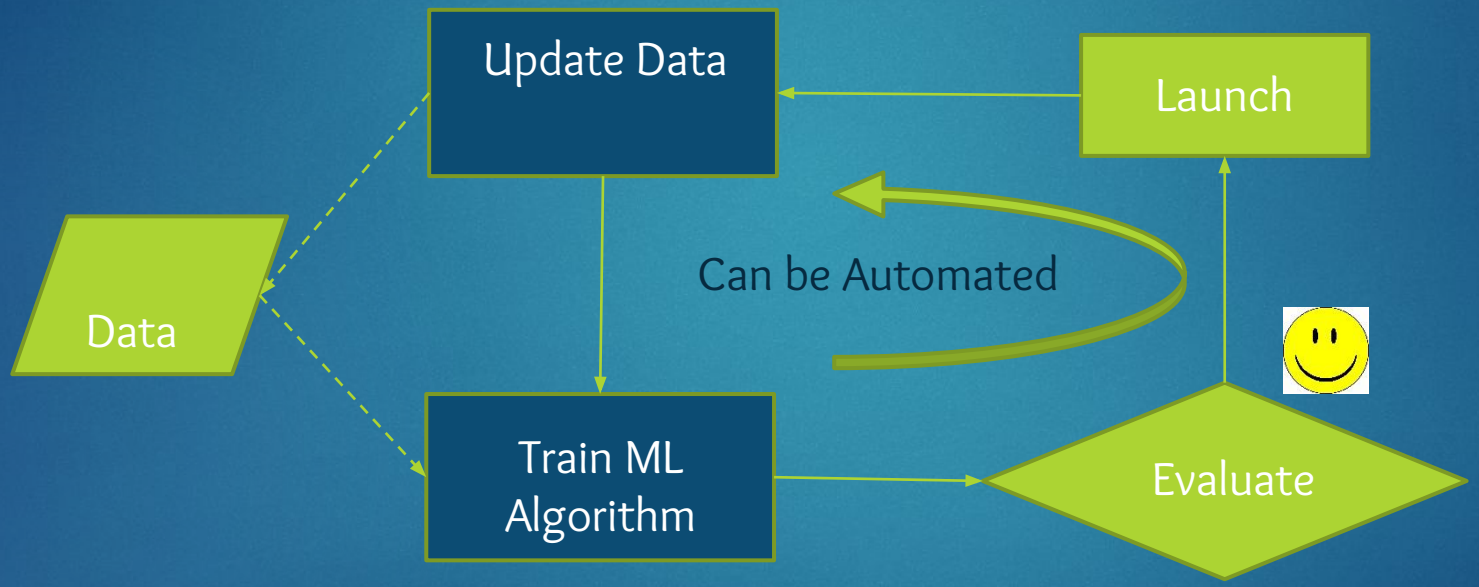
Why Use Machine Learning?

- Since the problem is not trivial, your program will likely become a long list of complex rules—pretty hard to maintain.
- Consider spam filter based on Machine Learning:



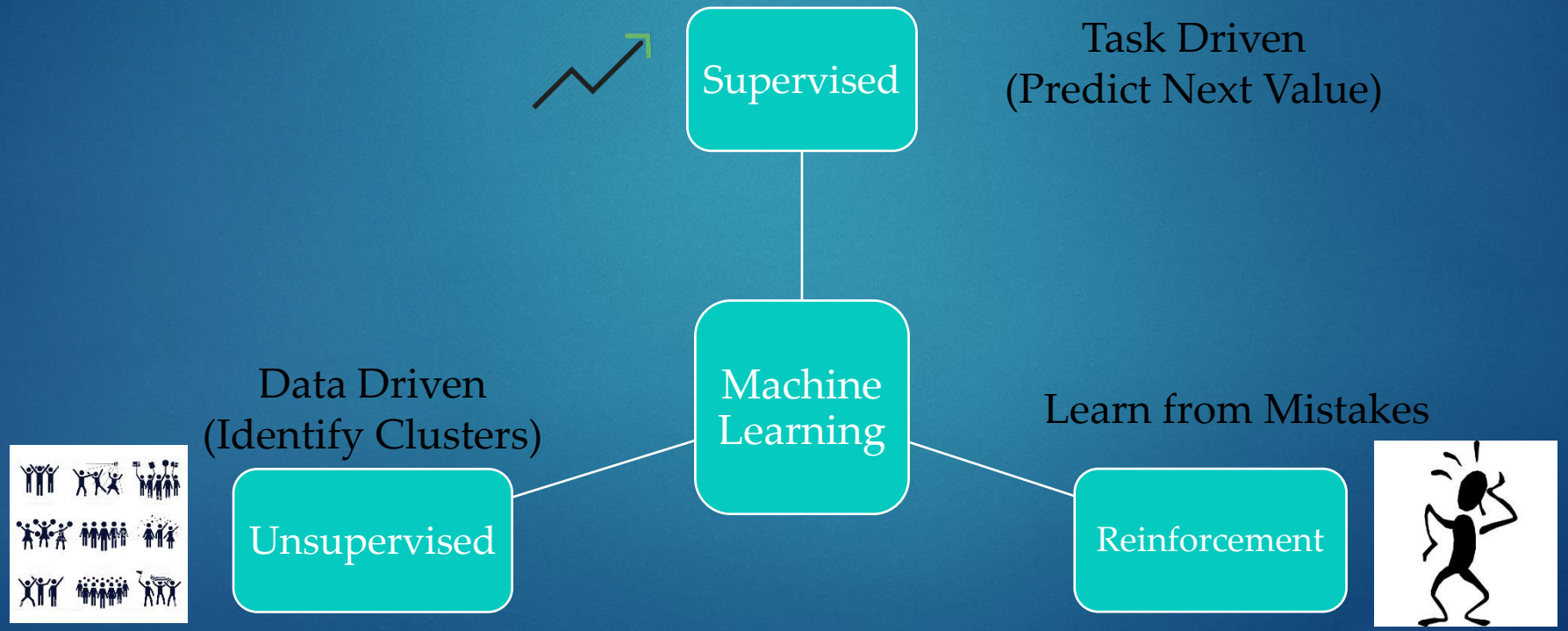
Why Use Machine Learning?

- Automatically adapts to change

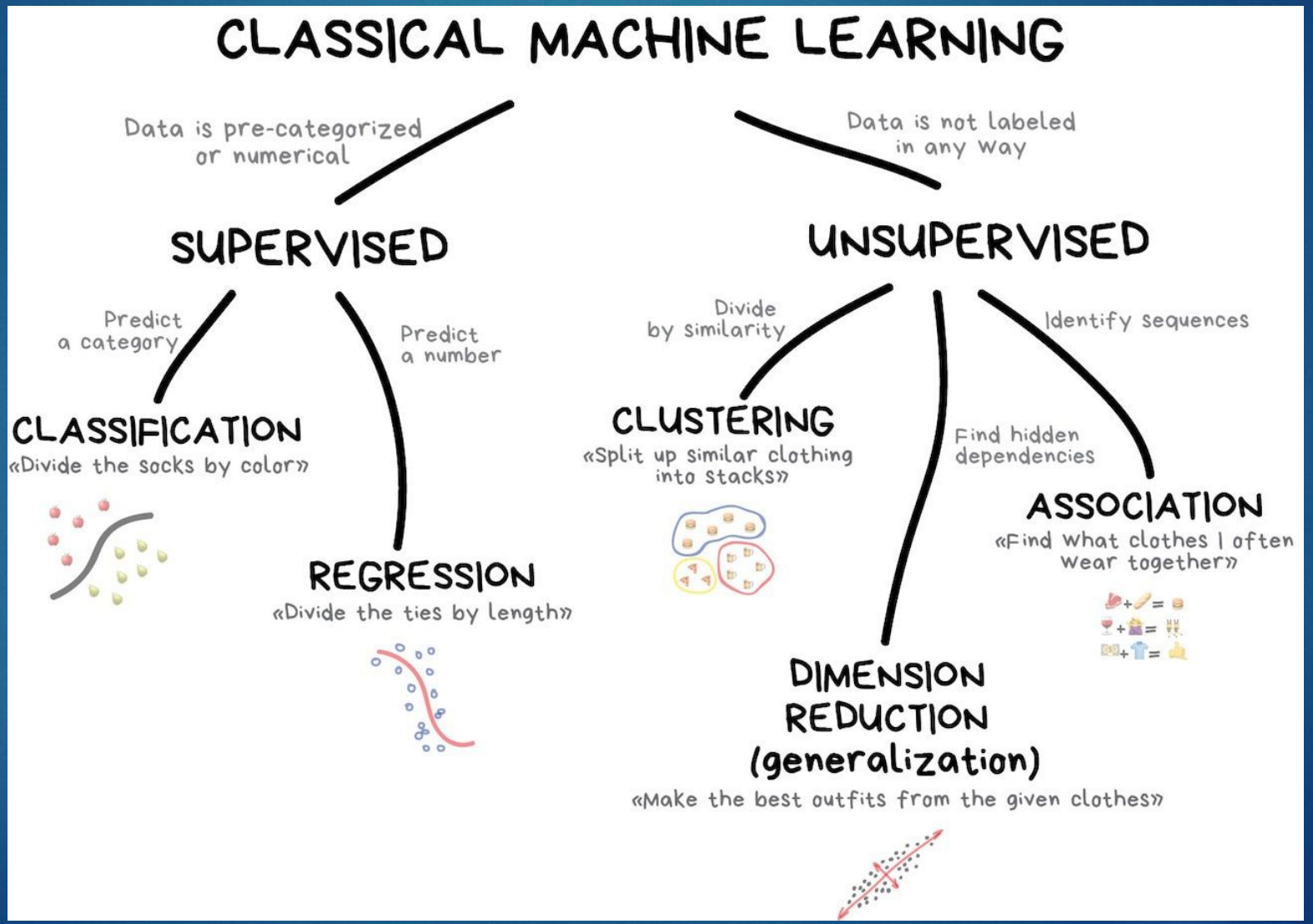


Types of Machine Learning

- Strictly speaking, we have three types of Machine Learning

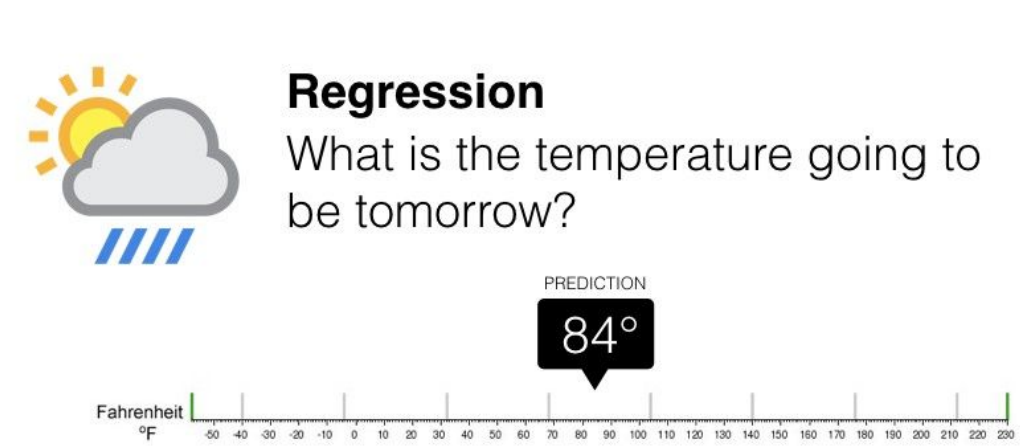


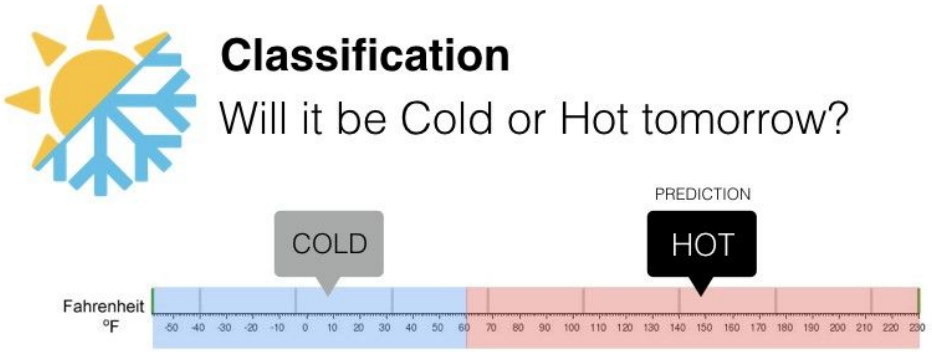
Classical Machine Learning



Classical Machine Learning

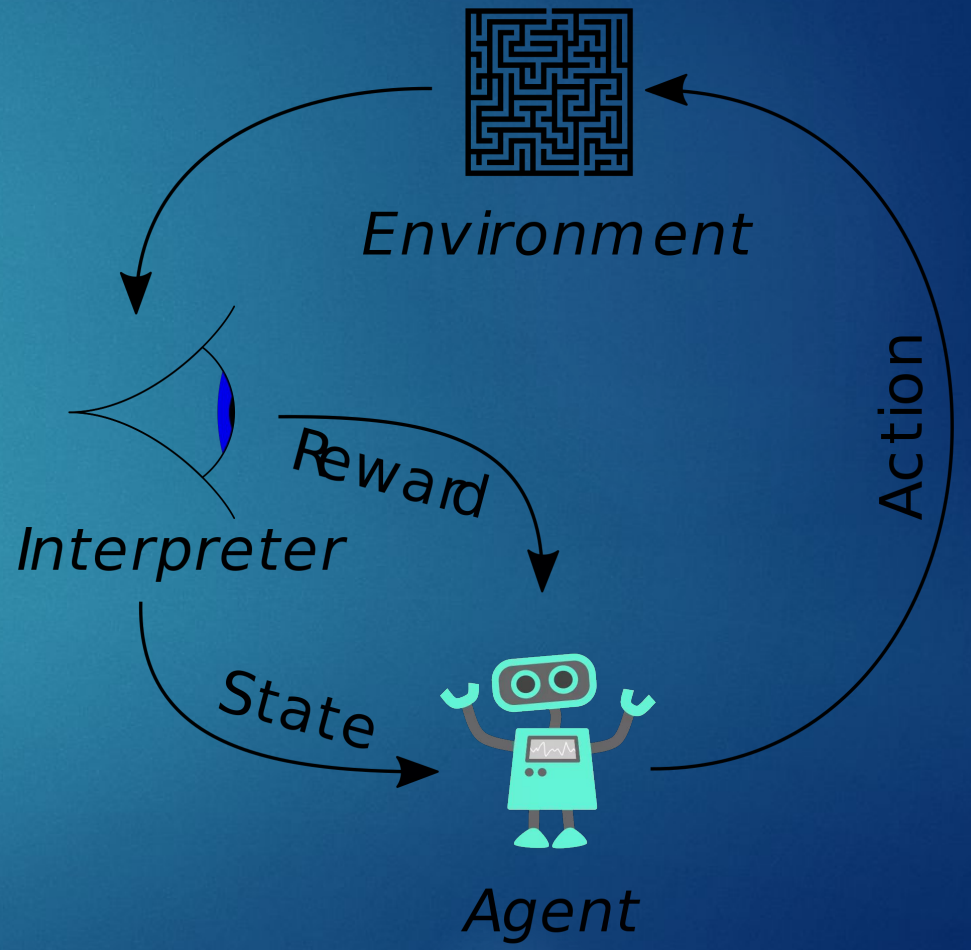
- Clearly, the machine will learn faster with a teacher, so it's more commonly used in real-life tasks.
- There are two types of such tasks:
  - Regression – prediction of a specific point on a numeric axis.
  - Classification – an object's category prediction.





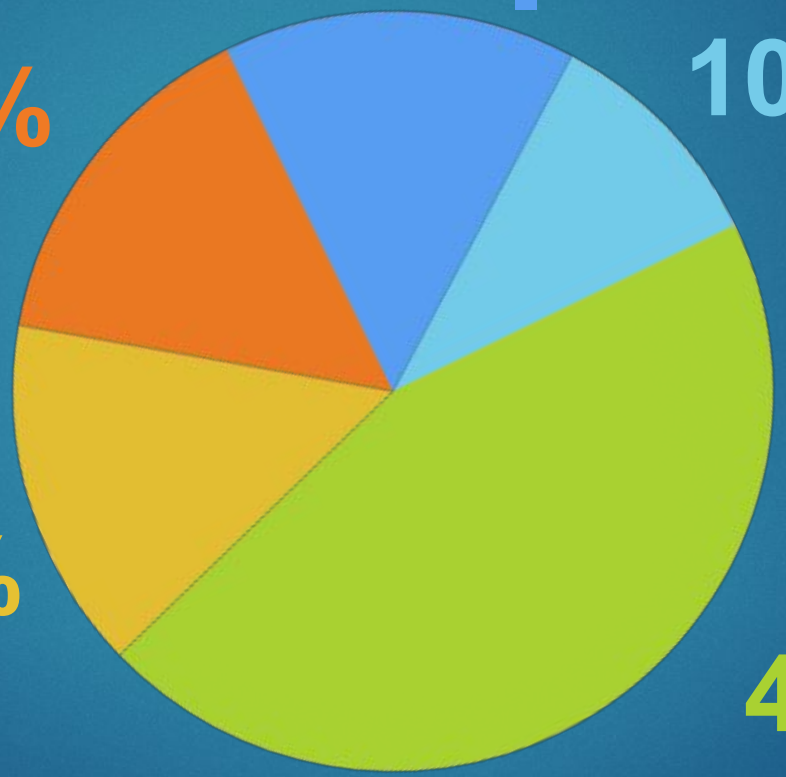
Reinforcement Learning

- *"Throw a robot into a maze and let it find an exit"*
- Nowadays used for:
  - Self-driving cars
  - Robot vacuums
  - Games
  - Automating trading
  - Enterprise resource management



What it takes to Learn Machine Learning?

- **Mathematics and Statistics**
  - Calculus
  - Differential equations
  - Mathematical statistics
  - Optimization
  - Regression and Time Series
  - Probability Distributions
  - Hypothesis Testing
  - Bayesian Modelling
  - Fitting of a distribution
- **Programming**
  - Object Oriented Programming
  - Classes
  - Objects
  - Constructors
  - Inheritance
  - Abstraction
  - Data Types
  - Conditional Operators
  - Control Statements


15%
Calculus
10%
Data Pre-Processing
Algorithms & Complexity
15%
Probability & Statistics
10%
45%
Linear Algebra

Happy Machine Learning!