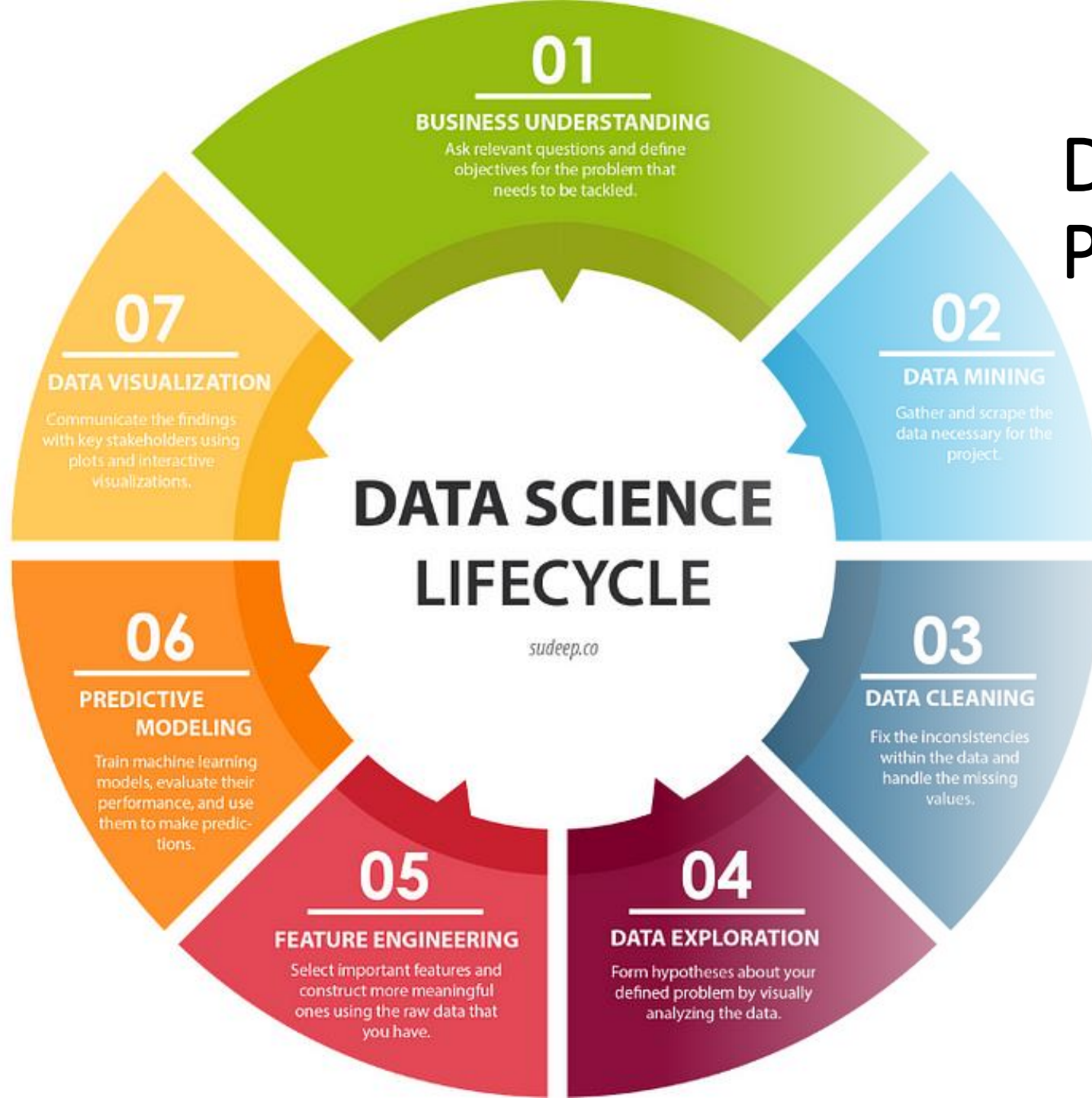


Data Science Pipeline



Identify your business needs and opportunities



Understanding your current situation and identifying the gaps, challenges, or opportunities you want to address with AI and ML is the first step in defining your AI and ML objectives.



To begin, ask yourself questions such as what are the main pain points or inefficiencies in your business processes or customer experience?



What are the key drivers or trends that affect your industry or market? What are the best practices or benchmarks that you want to match or surpass?



What are the new products, services, or features that you want to offer or improve?

Example of problem and solution statements

Example of problem statement-
“Our contract review process is very slow because my team and I spend many hours going through various Statement of Work and previously signed contracts to look for terms that have been acceptable to us previously.

Example of a solution statement -
"leverage Generative AI to retrieve and synthesize answers across documents and contracts and help reduce the time it takes to review new contracts."

Data Collection and Mining



Sensor Data



Privacy and
confidentiality



Too expensive or
time consuming



Non-existence



Poor accuracy



Storing data



etc

Data Cleaning

Clear
formatting

Remove
irrelevant data

Remove
duplicates

Filter missing
values

Delete outliers

Convert data
type

Standardize
capitalization

Structural
consistency

Uniform
language

Validate the
data

Data Exploration

Data exploration is the process of analyzing datasets to find patterns and relationships. It is called **exploratory data analysis (EDA)**.

Exploring data can help you to develop hypotheses about how different variables are related.

It can help you to identify which variables are the most important in predicting a particular outcome.

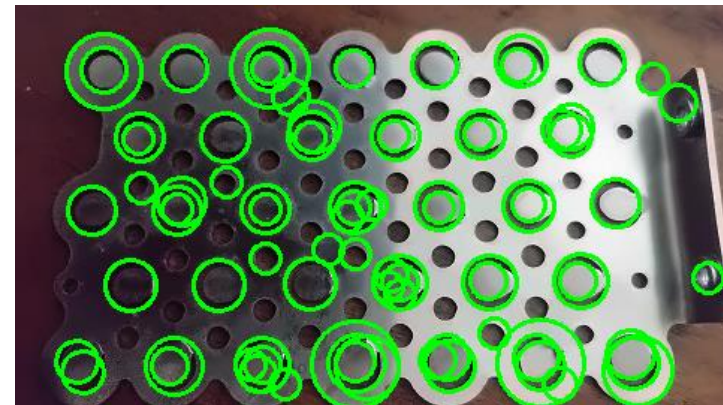
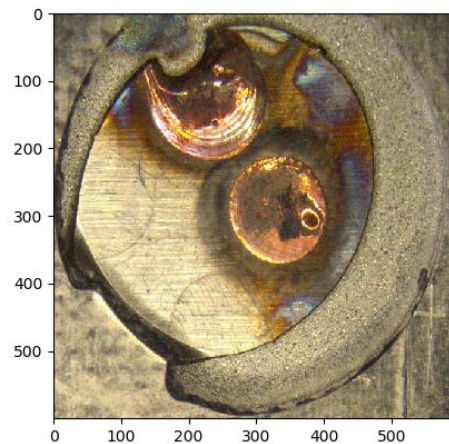
Exploring the data can help you to understand the data better and to develop intuition about how the data behaves.

<https://www.youtube.com/watch?v=h5NfWu5a4H0>

Laser Welding Car Battery

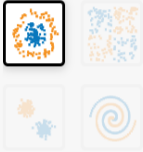


A Data Exploration Example



DATA

Which dataset do you want to use?



Ratio of training to test data: 80%

Noise: 0

Batch size: 14

REGENERATE

FEATURES

Which properties do you want to feed in?



+ - 2 HIDDEN LAYERS

+ -

3 neurons

+ -

3 neurons

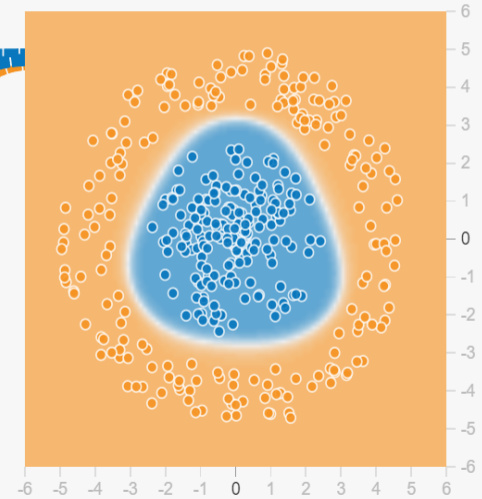
This is the output from one neuron. Hover to see it larger.

The outputs are mixed with varying weights, shown by the thickness of the lines.

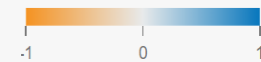
OUTPUT

Test loss 0.000

Training loss 0.000



Colors shows data, neuron and weight values



Feature Selection

- <https://playground.tensorflow.org>

Machine Learning Methods



- Neural Networks
- Deep Learning
- Genetic Algorithms/Evolutionary Computing
- Reinforcement Learning
- Association Rules
- Etc.

Visualization

