# Machine learning in a nutshell

Machine learning uses algorithms to identify patterns within data. Patterns that are used to create a **data model that can make predictions**. The **Azure for the Data Scientist course (DP-100)** focuses on creating and using machine learning models with **Azure Machine Learning**. To understand the purpose of the exercises, a simplified overview is provided here.
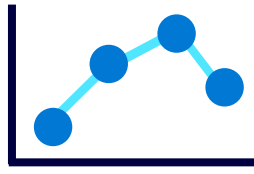
1. Define the problem
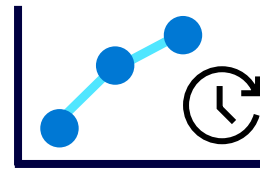Decide on what the model should predict and when it is successful.
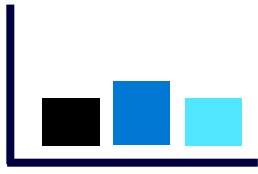
Classification
Predict a categorical value.
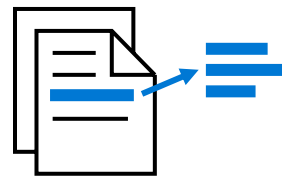
Regression
Predict a numerical value.

Time-series forecasting
Predict future numerical values based on time-series data.

Computer Vision
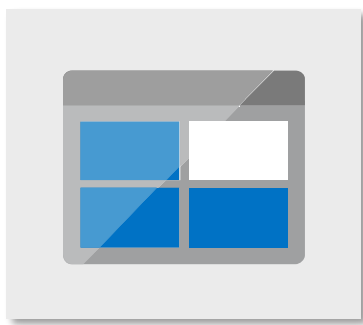Classify images or detect objects in images.

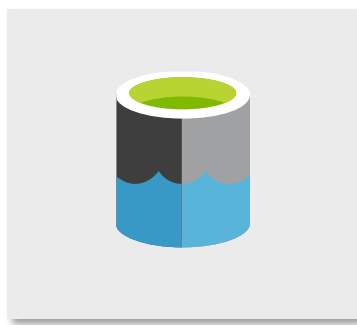**Natural Language Processing**
Extract insights from text.

2. Get the data
Find data sources and get access. Azure Machine Learning connects seamlessly with the three most used Azure storages for data science.
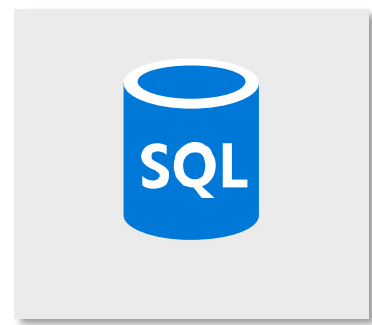
Azure Blob Storage
Object cloud storage. Uses flat namespace to store unstructured data.

Azure Data Lake Gen2
Unlimited object cloud storage. Uses hierarchical namespace for granular access control.

Azure SQL Database
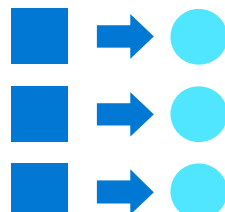Relational cloud database. Used for tabular and transactional data.

3. Prepare the data
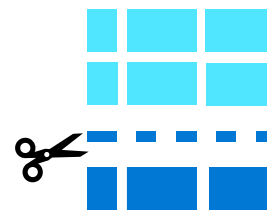Explore the data. Clean and transform the data based on the model's requirements.

Exploratory data analysis (EDA): Analyse your data, get summary statistics, and understand possible correlations between variables.

Feature engineering: Transform the data to create features that will help the model to predict the target value.

Create validation set: Split the data into training and validation or test dataset to evaluate the model.
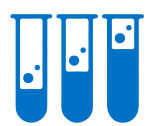
4. Train the model
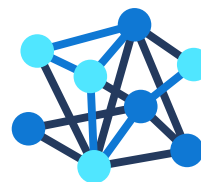Choose algorithm and hyperparameters based on trial and error.

Data
Includes features (what influences the to be predicted value) and the target value (if it exists).

Algorithm
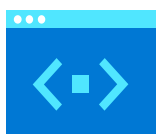Based on the task (e.g. classification), different algorithms and hyperparameters can be tried.

Model
Often stored as a binary file (e.g. pickle file). Use on new data with same features to predict the target value.
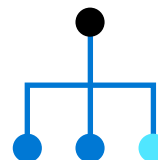
5. Integrate the model
Use endpoint to generate predictions.

Real-time predictions
Create light-weight app to predict target value in real-time per new data measurement.

Batch predictions
Create pipeline to predict target value on new set of data measurements.

6. Monitor the model
Track the model's performance.

Data drift
When new data differs significantly from training dataset.

Evaluation metrics
Keep track of the model's performance. When the model's predictions are increasingly incorrect.

Retrain the model