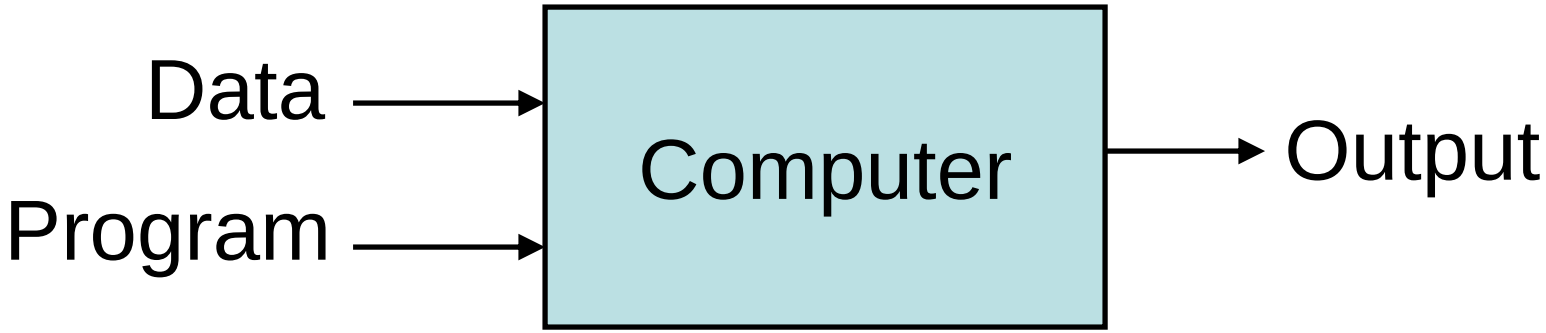
Machine Learning

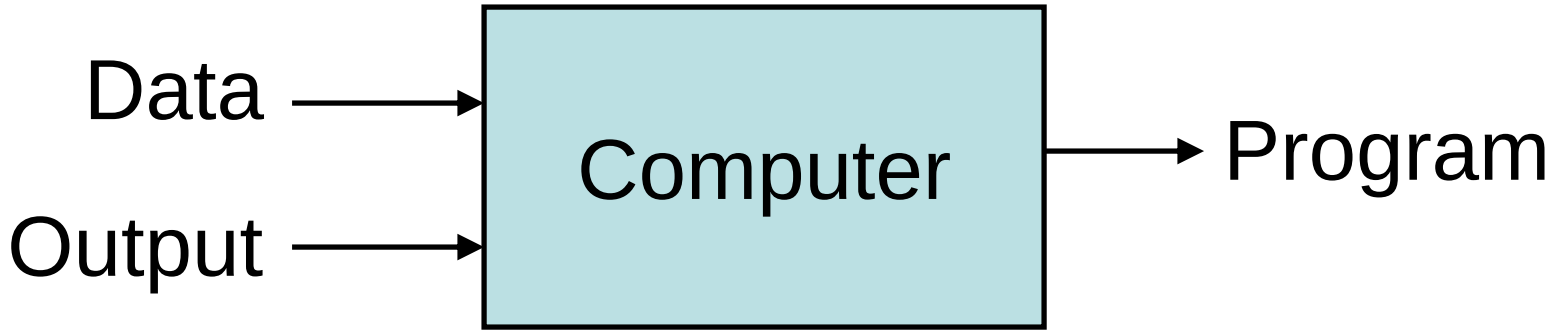
What Is Machine Learning?

- A subfield of computer science and artificial intelligence (AI) that focuses on the design of systems that can learn from and make decisions and predictions based on data.
- Machine learning enables computers to act and make data-driven decisions rather than being explicitly **programmed** to carry out a certain task

Traditional Programming



Machine Learning



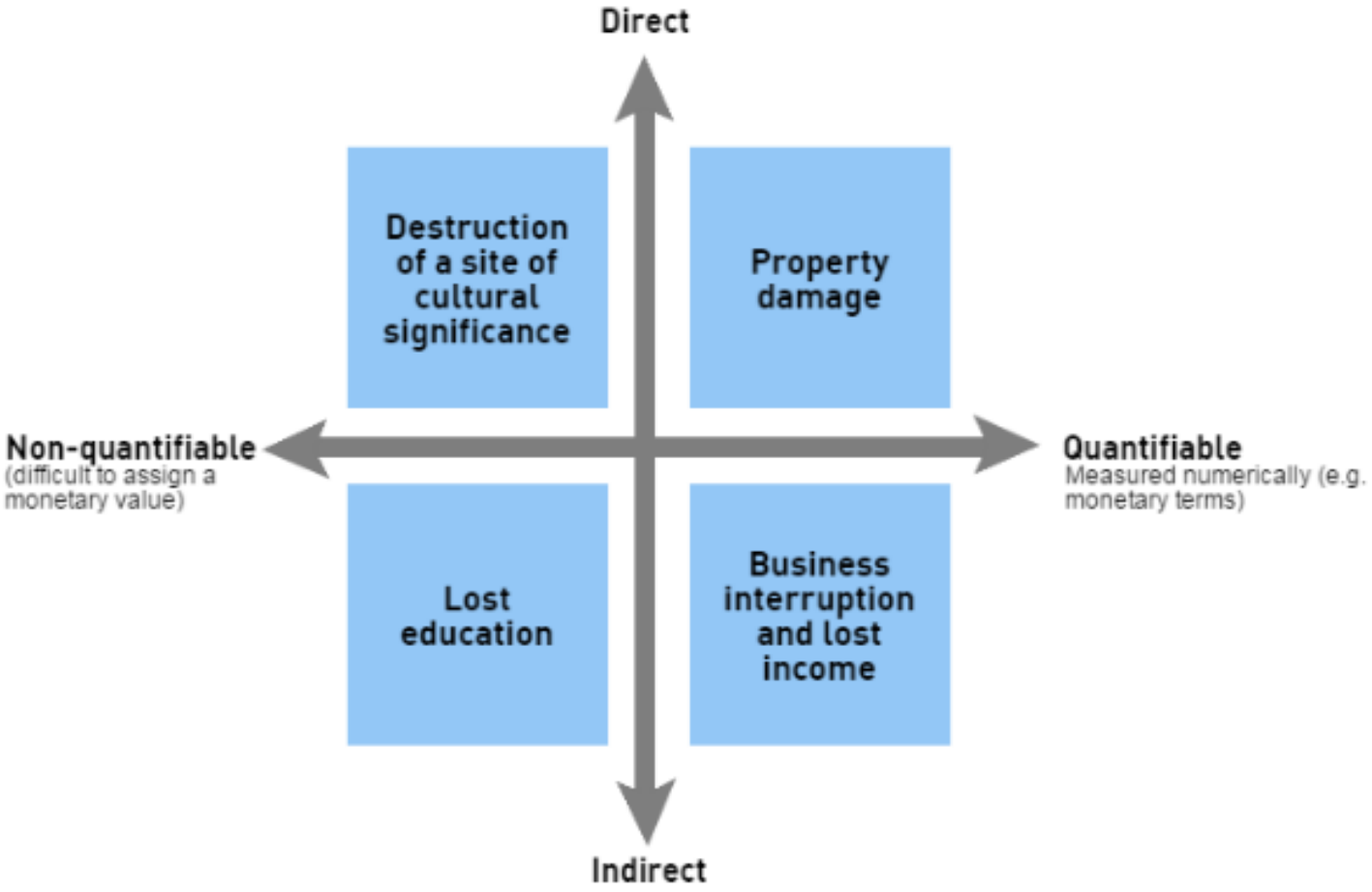
Identification of unwanted spam messages in e-mail

Segmentation of customer behavior for targeted advertising

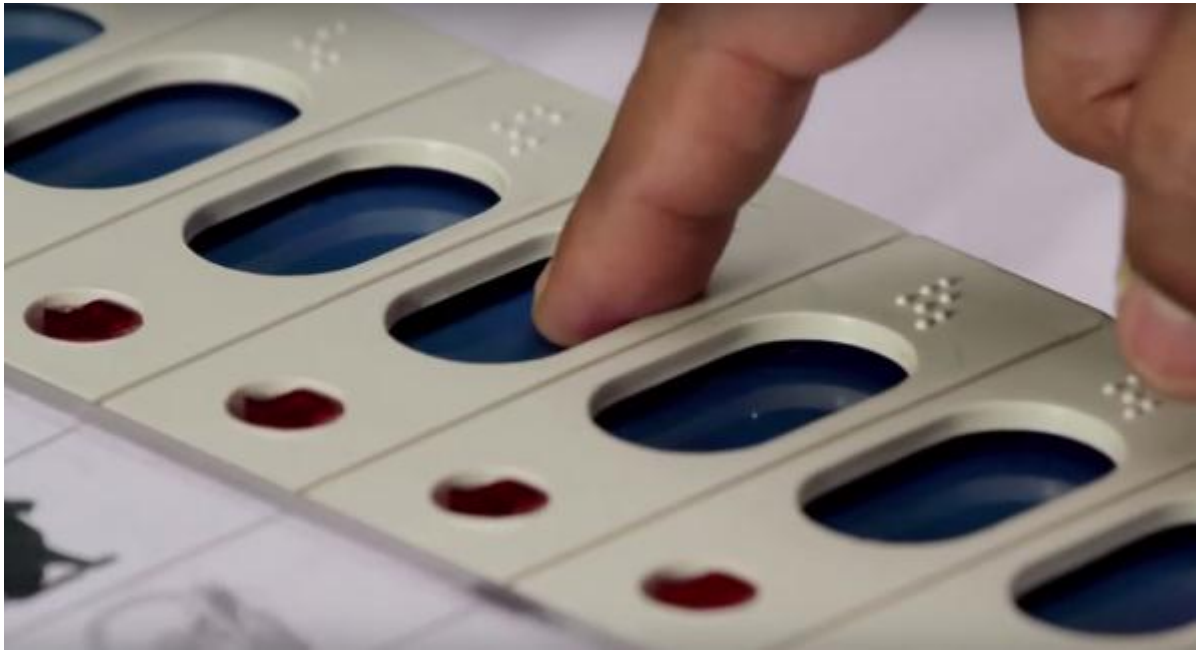
Forecasts of weather behavior and long-term climate changes

Reduction of fraudulent credit card transactions

Estimation of financial damage of natural disasters

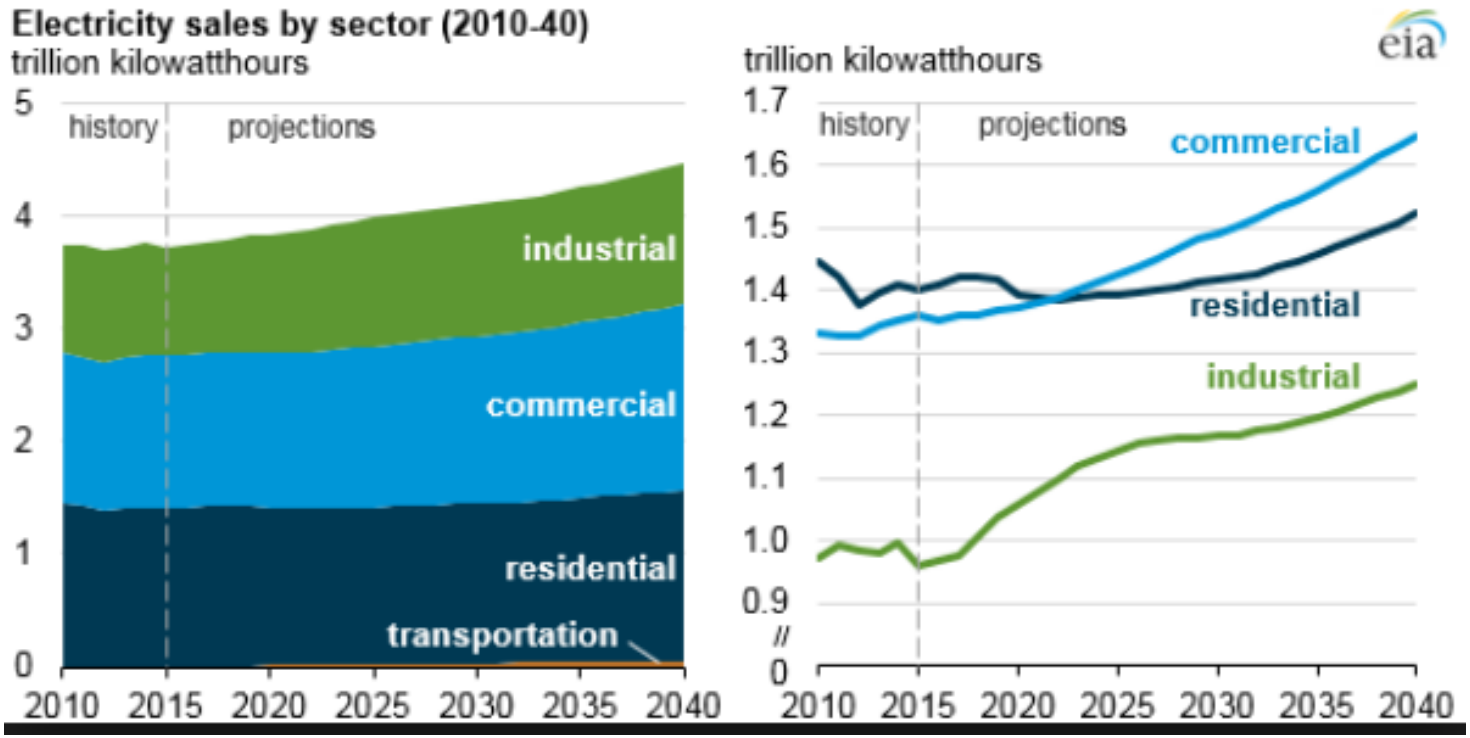


Prediction of popular election outcomes



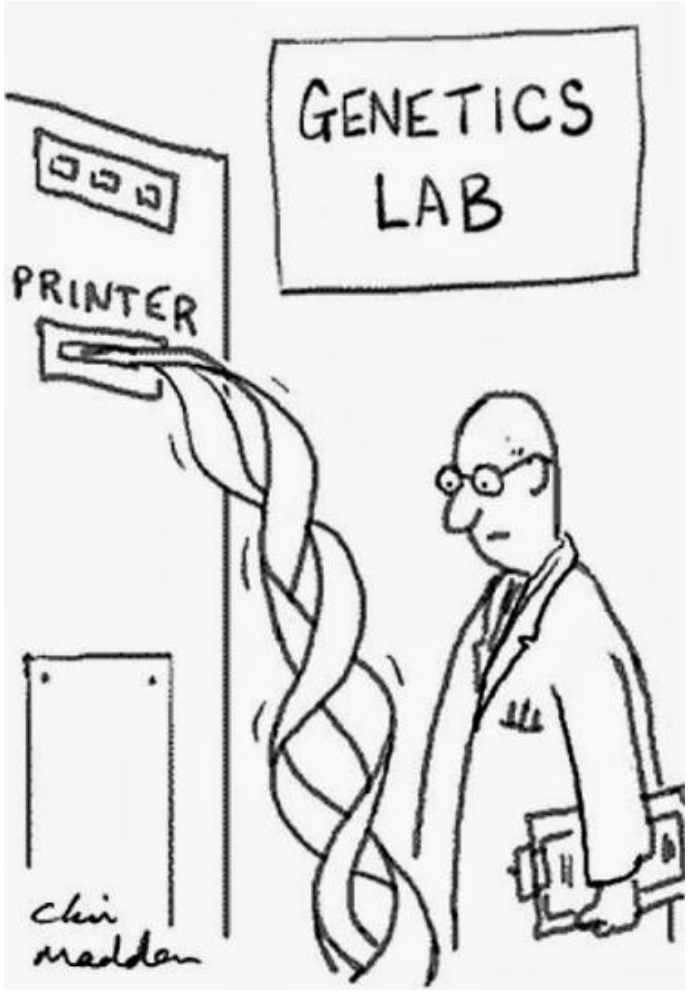
Development of algorithms for auto-piloting drones

Optimization of energy use in homes and office buildings



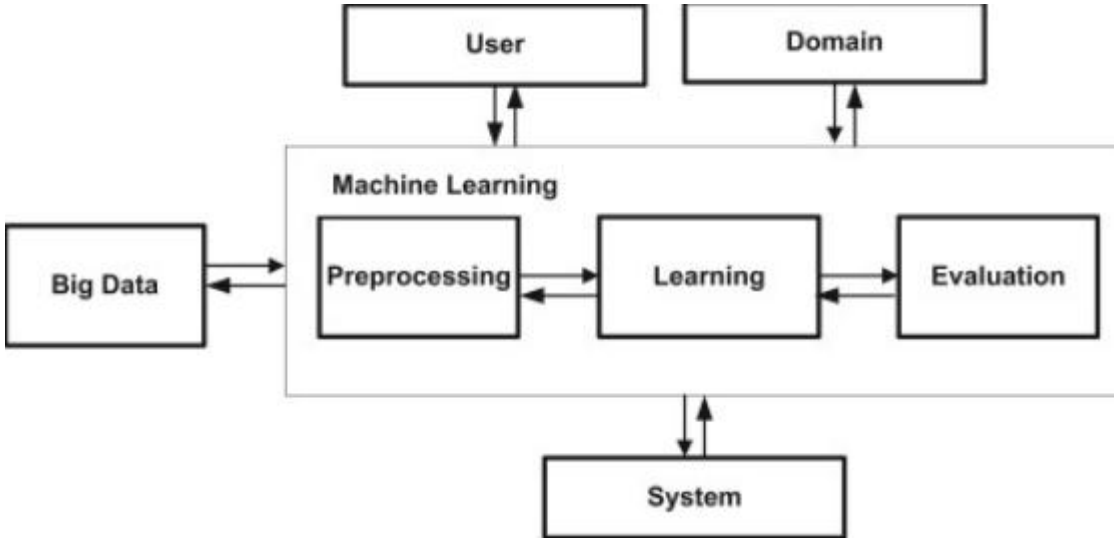
Projection of areas where criminal activity is most likely

Discovery of genetic sequences linked to diseases

- 

[Your favorite area]

ML Model


User
Domain
Machine Learning
Big Data
Preprocessing
Learning
Evaluation
System

ML Algorithms

- Tens of thousands of machine learning algorithms
- Hundreds new every year
- Every machine learning algorithm has three components:
  - **Representation**
  - **Evaluation**
  - **Optimization**

Representation

- Linear Regression
- Nearest Neighbour
- Naive Bayes
- Decision Trees
- Neural Networks
- Support Vector Machines
- Association Rules
- K-means clustering

Evaluation

- Accuracy
- Confusion Matrix
- Precision and recall
- F1 Score
- RMSE
- Loss Function
- Margin
- Etc.

Optimization

- Combinatorial optimization
  - E.g.: Greedy search
- Convex optimization
  - E.g.: Gradient descent
- Constrained optimization
  - E.g.: Linear programming

Types of Learning

- **Supervised (inductive) learning**
  - Training data includes desired outputs
- **Unsupervised learning**
  - Training data does not include desired outputs
- **Semi-supervised learning**
  - Training data includes a few desired outputs
- **Reinforcement learning**
  - Rewards from sequence of actions

ML Algorithms Tasks

(Algorithms Grouped by its Task)

- **Classification (Binary and Multiclass)**
- **Regression (Numeric Prediction)**
- **Clustering (Identifying Clusters)**
- **Finding Associations (Pattern Detection)**

- **Dimensionality Reduction**
- **Feature Selection**

Feature Selection and Dimensionality Reduction

Both methods seek to reduce the number of attributes in the dataset

- Dimensionality reduction method do so by creating new combinations of attributes,
- Feature selection methods include and exclude attributes present in the data without changing them.

Supervised Learning Algorithms and their category

- Nearest Neighbour -Classification
- Naive Bayes -Classification
- Decision Trees - Classification
- Linear Regression -Numeric prediction
- Regression Trees -Numeric prediction
- Neural Networks -Dual use
- Support Vector Machines -Dual use

Unsupervised Learning

- A priori algorithm -Association rule Mining/Pattern detection
- k-means clustering - Clustering

Supervised Learning

- **Given** examples of a function *(X, F(X))*
- **Predict** function *F(X)* for new examples *X*
  - Discrete *F(X)*: Classification
  - Continuous *F(X)*: Regression

Question

- Identify 5 application areas of ML and say the task to be done by the algorithm (classification, regression, association mapping etc) and the method of learning required/possible to do the task (supervised/unsupervised)

Machine learning algorithm deployment steps:

1. Data collection:
2. Data exploration and preparation:
3. Model training
4. Model evaluation
5. Model improvement

A Few Quotes

- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet”
  (Tony Tether, Director, DARPA)
- Machine learning is the hot new thing”
  (John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, Dir. Research, Yahoo)
- “Machine learning is going to result in a real revolution”
  (Greg Papadopoulos, CTO, Sun)
- “Machine learning is today’s discontinuity”
  (Jerry Yang, CEO, Yahoo)

Sample Applications

- Identification of unwanted spam messages in e-mail
- Segmentation of customer behavior for targeted advertising
- Forecasts of weather behavior and long-term climate changes
- Reduction of fraudulent credit card transactions
- Actuarial estimates of financial damage of natural disasters
- Prediction of popular election outcomes
- Development of algorithms for auto-piloting drones
- Optimization of energy use in homes and office buildings
- Projection of areas where criminal activity is most likely
- Discovery of genetic sequences linked to diseases
- [Your favorite area]