Introduction to Machine Learning

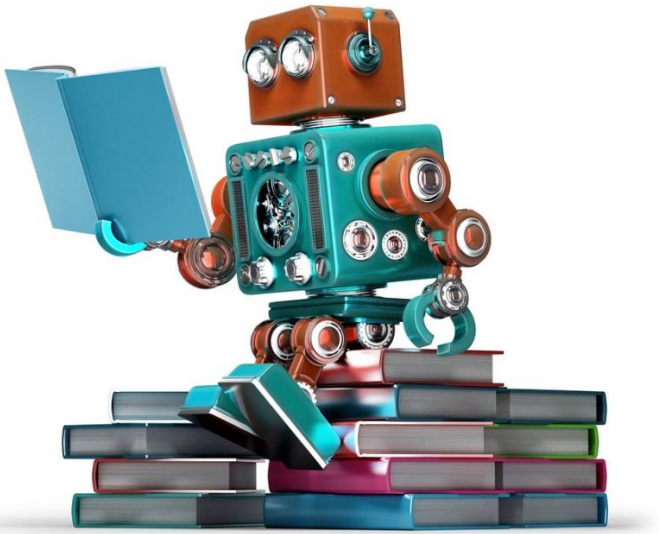
Machine Learning

Machine learning is the subfield of computer science that, according to Arthur Samuel, gives "computers the ability to learn without being explicitly programmed." Samuel, an American pioneer in the field of computer gaming and artificial intelligence, coined the term "machine learning" in 1959 while at IBM.

Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome. Machine learning is so pervasive today that you probably use it dozens of times a day without knowing it.

What is Machine Learning ?

Machine Learning is a field of study that gives computers the ability to learn without being explicitly programmed



What is Machine Learning?

The ability to perform a task in a situation which has never been encountered before (Learning = Generalization)

Automating automation

Getting computers to program themselves

Writing software is the bottleneck

Let the data do the work instead!

A short story

Samuel's claim to fame was that back in the 1950, he wrote a checkers playing program and the amazing thing about this checkers playing program was that Arthur Samuel himself wasn't a very good checkers player. But what he did was he had to programmed maybe tens of thousands of games against himself, and by watching what sorts of board positions tended to lead to wins and what sort of board positions tended to lead to losses, the checkers playing program learned over time what are good board positions and what are bad board positions. And eventually learn to play checkers better than the Arthur Samuel himself was able to. This was a remarkable result.

Arthur Samuel himself turns out not to be a very good checkers player. But because a computer has the patience to play tens of thousands of games against itself, no human has the patience to play that many games. By doing this, a computer was able to get so much checkers playing experience that it eventually became a better checkers player than Arthur himself.

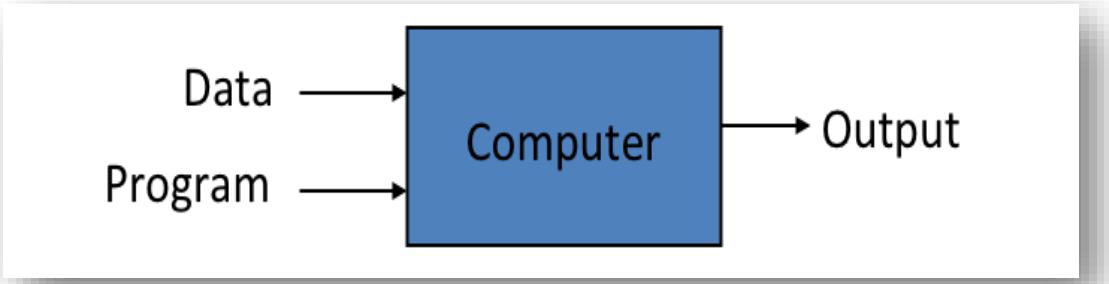
What is Machine Learning

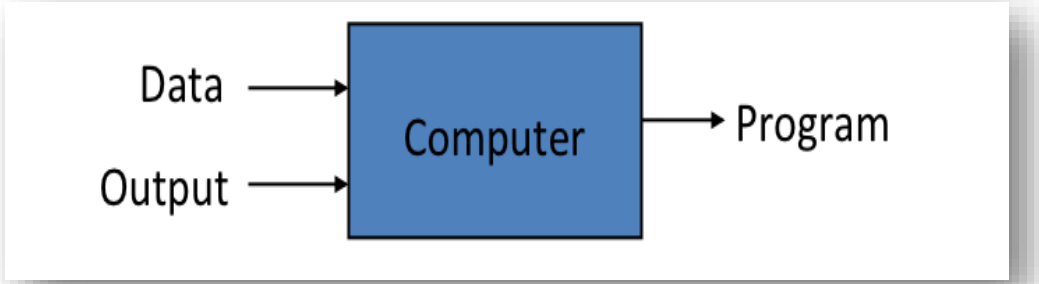
- “A computer program is said to learn from experience E with some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.” -Tom M. Mitchell Consider **playing checkers.**
- E = the experience of playing many games of checkers
- T = the task of playing checkers.
- P = the probability that the program will win the next game

- classification problems where the goal is to categorize objects into a fixed set of categories.
- **Face detection**: Identify faces in images (or indicate if a face is present).
- **Email filtering**: Classify emails into spam and not-spam.
- **Medical diagnosis**: Diagnose a patient as a sufferer or non-sufferer of some disease.
- **Weather prediction**: Predict, for instance, whether or not it will rain tomorrow.

- Facial recognition technology allows social media platforms to help users tag and share photos of friends.
- Optical character recognition (OCR) technology converts images of text into movable type
- Recommendation engines, powered by machine learning, suggest what movies or television shows to watch next based on user preferences.
- Self-driving cars that rely on machine learning to navigate

Traditional Programming Vs Machine Learning





Why Machine Learning?

It has the potential to unlock the value of corporate and customer data.

It can assist in strategic decision-making, which can keep a company ahead of the competition

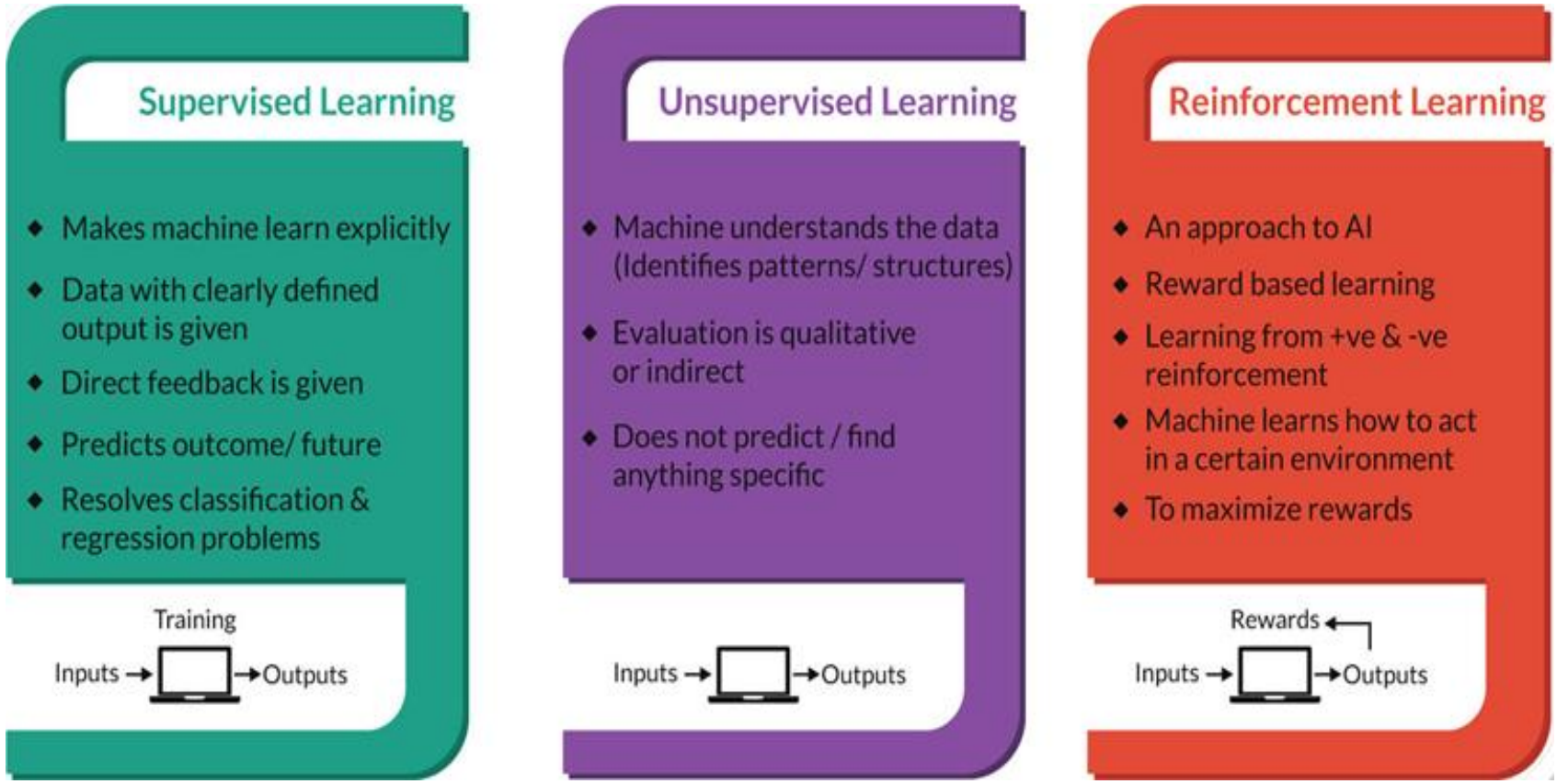
It is going to change the ways humans operate more than any technology that’s ever existed.

Types of Machine Learning Algorithms

Machine Learning Algorithms are divided into two types:

- Supervised Machine Learning Algorithms
- Unsupervised Machine Learning Algorithms

Types of Learning

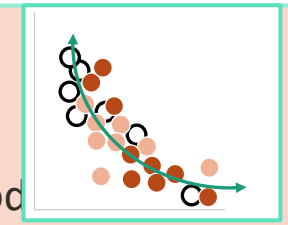


Types of Machine Learning Algorithms: Supervised

- Supervised learning algorithms involve direct supervision of the operation.
- It is a spoon-fed version of machine learning:
  - Select what kind of information output (samples) to “feed” the algorithm.
  - Select what kind of results are desired (for example, “yes/no” or “true/false”).
- Supervised machine learning includes two major processes:
  - Classification
  - Regression

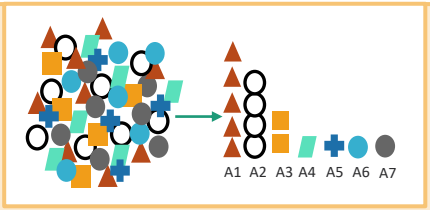
Machine Learning Algorithms

Regression



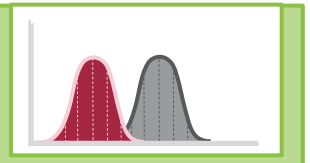
- Linear Model
- Generalized Linear Mod
- Support Vector Machine (SVM)
- Stepwise Linear Regression
- Neural Network
- LASSO *

Attribute Importance



- Minimum Description Length
- Principal Comp Analysis (PCA)
- Unsupervised Pair-Wise KL Div
- CUR Decomposition for Row & Al

Statistical Functions



- Basic Statistics: min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.

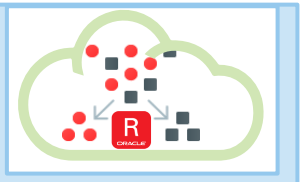
R Packages

- CRAN R Algorithm Packages Through Embedded R Execution
- Spark MLlib Algorithm Integration

Exportable ML Models

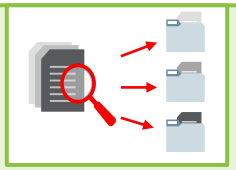
- REST APIs for Deployment

Feature Extraction



- Principal Comp Analysis (PCA)
- Non-Negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

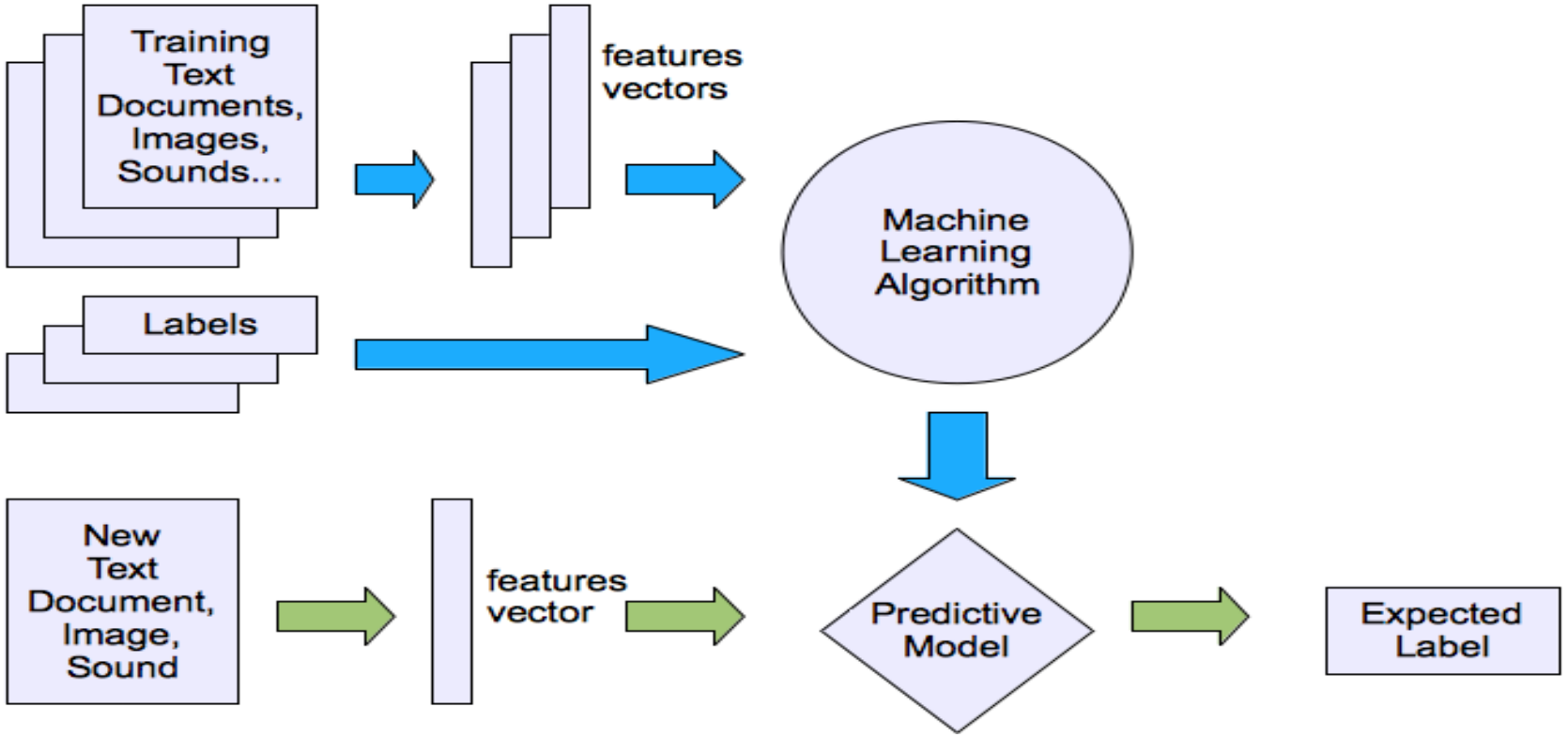
Text Mining Support



- Algorithms Support Text
- Tokenization and Theme Extraction
- Explicit Semantic Analysis (ESA) for Document Similarity

Machine learning structure

- Supervised learning

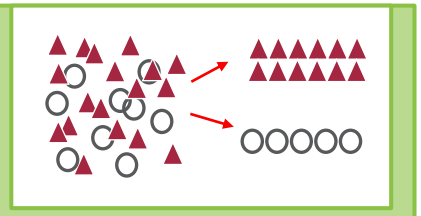


Types of Machine Learning Algorithms: Unsupervised

- Unsupervised learning does not involve direct control of the developer.
- Unsupervised learning is non-directed.
- There is no distinction between dependent and independent attributes.
- There is no previously-known result to guide the algorithm in building the model.

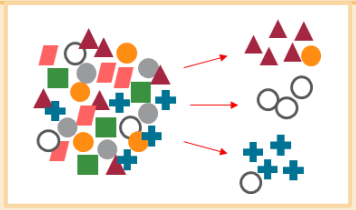
Machine Learning Algorithms

Classification



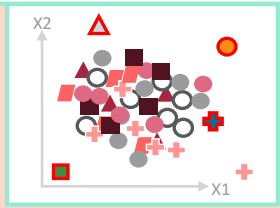
- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine

Clustering



- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)

Anomaly Detection



- One-Class SVM

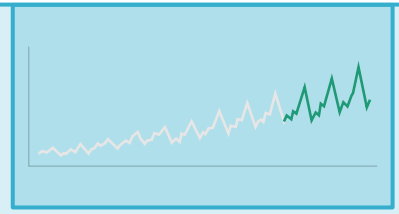
Association Rules

- A Priori/Market Basket

SQL Analytics

- SQL Windows, SQL Patte
  SQL Aggregates

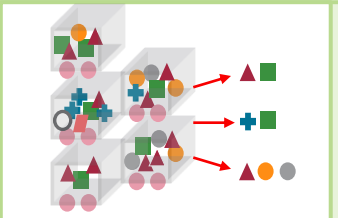
Time Series



- State-of-the-art forecasting using Exponential Smoothing
- Includes all popular models e.g. Holt-Winters with trends, seasons, irregularity, missing data

Predictive Queries

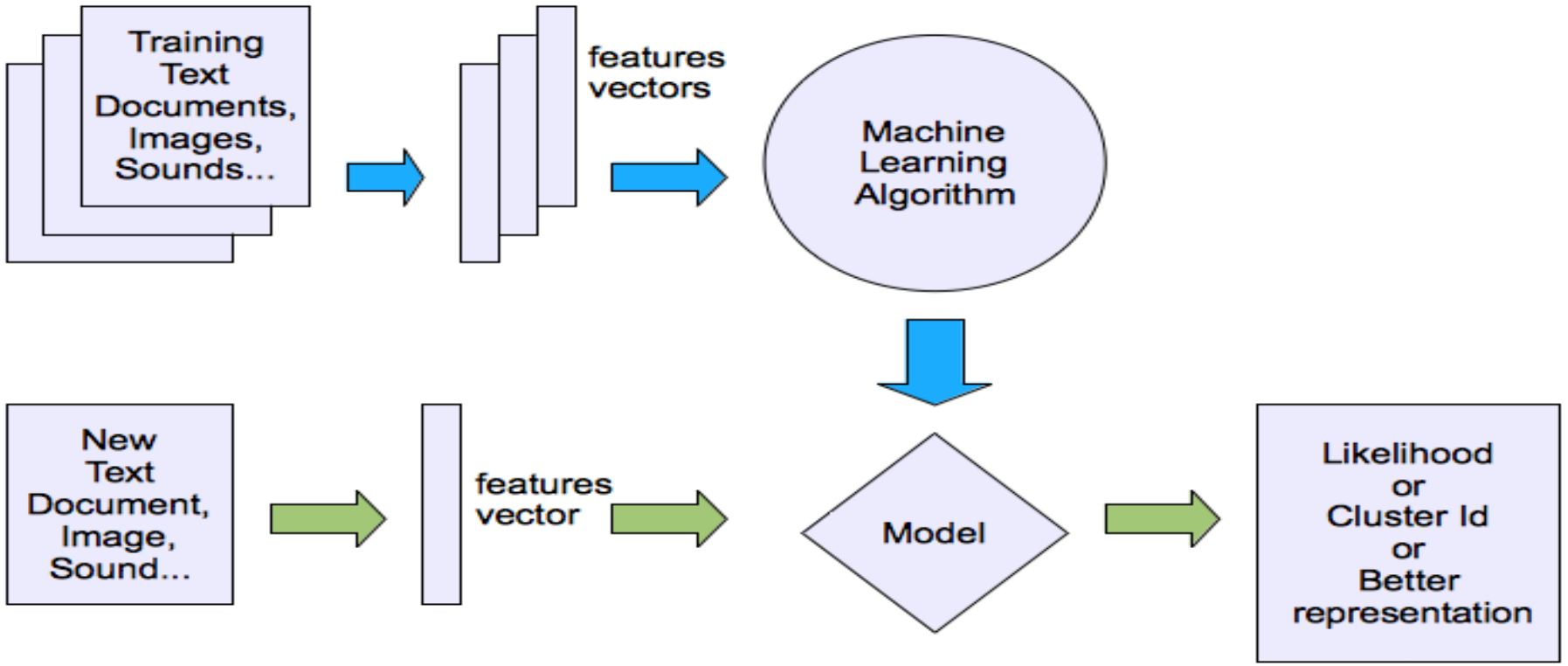
- Predict, Cluster, Detect, Features



- Unsupervised learning algorithms apply the following techniques to describe data:
- **Clustering:** It is an exploration of data to segment it into meaningful groups (i.e., clusters) based on internal patterns without prior knowledge of group credentials.
- The credentials are defined by the similarity of individual data objects and aspects of its dissimilarity from the rest (which can also be used to detect anomalies).
- **Dimensionality reduction:** There is a lot of noise in the incoming data. Machine learning algorithms use dimensionality reduction to remove this noise while distilling the relevant information.
- The most widely used algorithms are:
  - k-means clustering
  - PCA (Principal Component Analysis)
  - Association rule

Machine learning structure

- Unsupervised learning



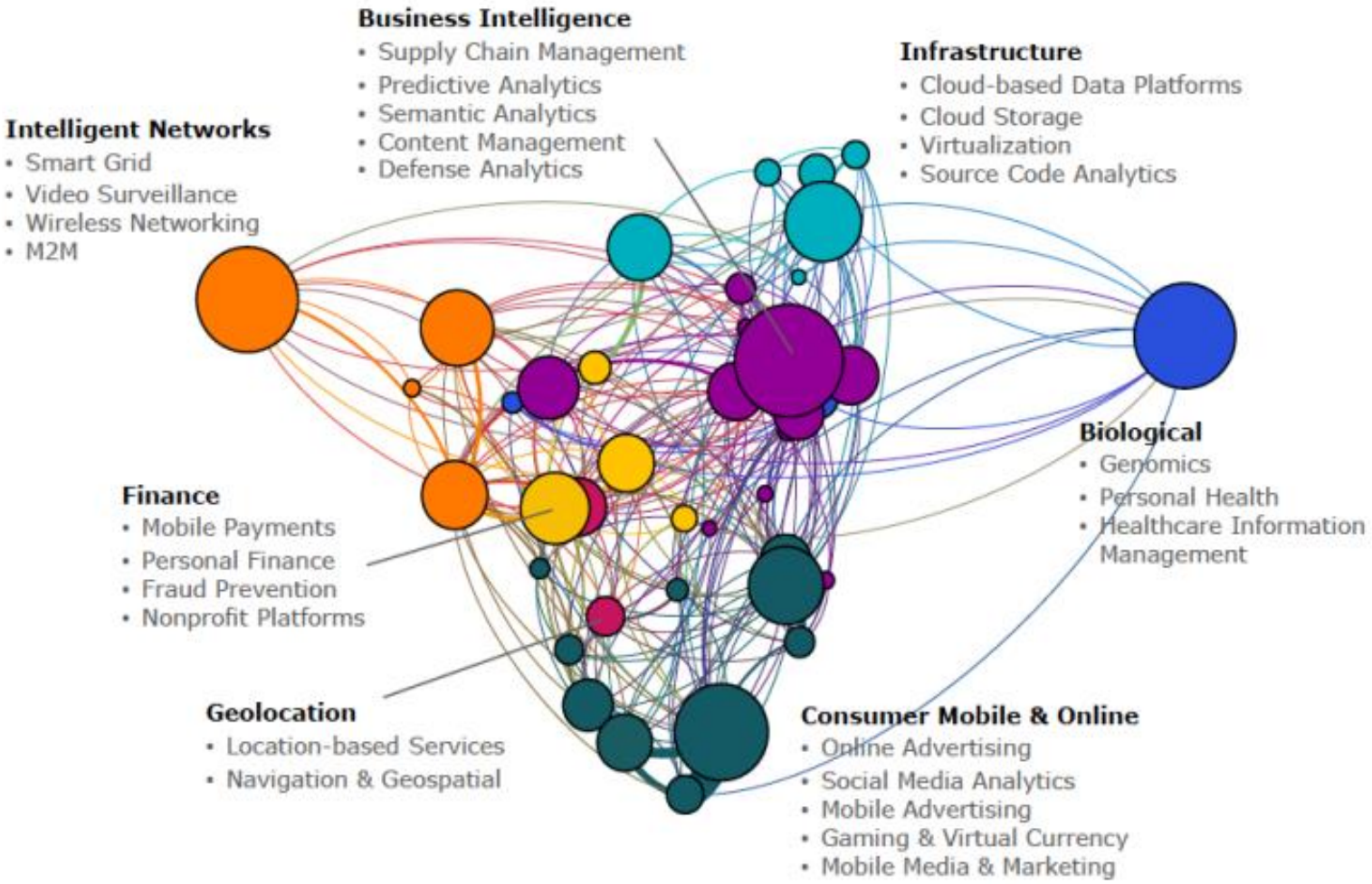
Algorithms

- The success of machine learning system also depends on the algorithms.
- The algorithms control the search to find and build the knowledge structures.
- The learning algorithms should extract useful information from training examples.

Technologies Used

Machine Learning Software’s
Tensor Flow
PyTorch
Scikit-Learn
Keras
XGBoost
Apache Spark Mllib
Microsoft Azure Machine Learning
RapidMiner

Applications Of Datascience



Top Data Science Trends For 2024

- Augmented Analytics
- Responsible AI
- Edge Computing for Data Science
- Quantum Computing Integration
- Continuous Learning Models
- Natural Language Processing (NLP) Advancements
- Federated Learning
- Blockchain in Data Science

The data science process