Machine Learning

The machine learning is the subset of AI that enables computer systems to automatically learn & improve from experience without being explicitly programmed. It involves developing algorithms that can identify patterns in data, make decisions, and adapt over time with minimal human intervention

Machine learning Paradigms:

- Supervised learning models train on labelled data
- Unsupervised learning works without labels
- Semi-supervised learning works iteratively
- Reinforcement learning learns from feedback

Machine Learning techniques:

- Classification Predicts class or category of a case
- Regression / estimation Predicts continuous values
- Clustering groups similar data
- Association finds items or events that cooccur

- Anomaly detection discovers abnormal and unusual case
- Sequence mining Predicts next event
- Dimension reduction reduce size of data
- Recommendation systems associate people's Preferences

⊙ Process Lifecycle

Problem Definition → Data Collection → Data Preparation → Model Development and Evaluation → Model Deployment → (back to Problem Definition)

⊙ Machine learning tools:
- Provide functionalities for ML pipelines

○ Include modules for:
- Data preprocessing and building
- Evaluating
- Optimizing
- Implementing ML models

Examples:
- Pandas: Data manipulation and analysis
- Scikit-learn: Supervised and Unsupervised Learning algorithms for linear regression

- Commonly Used Language in ML
  - Python
  - R
  - Julia
  - Scala
  - Java
  - JavaScript

- Data Processing and analytics
  - PostgreSQL
  - Hadoop
  - Spark
  - Apache Kafk

- Data visualization:
  - Matplotlib
  - Seaborn
  - ggplot 2
  - Tableau

- Machine learning:
  - Numpy
  - Pandas
  - Scipy
  - Scikit learn

- Deep learning:
  - TensorFlow
  - Keras
  - Theano
  - PyTorch

- Computer vision:
  - Open CV
  - sckit-image
  - Touch vision
- NLP:
  - NLTK
  - TextBlob
  - Stanza
- Generative AI
  - Hugging face transformation
  - ChatGPT
  - DALL-E
  - PyTorch

⊙ Machine Learning Ecosystems.

1. Data Collection
2. Preprocessing —— Machine learning tools
3. Model training
4. Model evaluation
5. Model deployment
6. Monitoring

- ML Ecosystem
  - Interconnected tools, frameworks, libraries, platforms, and processes
  - Supports developing, deploying, and managing ML models

- o Scikit-learn
  - Free ML Library for Python
  - Classification, regression, clustering and dimensionality reduction algorithm
  - Designed to work with NumPy and Scipy
  - Excellent documentation and community support
  - Constantly evolving
  - Enables easy implementation of ML models

- o Machine learning Pipeline tasks

3. Model setup and fitting	4. Hyperparameter tuning with Cross-Validation
2. Train or test splitting	5. Prediction
	Implemented in Scikit-learn
1. Data preprocessing task	6. Evaluation
	7. Exporting the model to be used in Production

- ⊙ Module 2
  - − Linear Regression
    - Supervised Learning model
    - Models a relationship between a continuous target variable and explanatory features
  - o Types of regression
    - Simple regression (Simple linear, Simple Nonlinear)
    - Multiple regression (Multiple linear, Multiple Nonlinear)