Core AI & Modeling Concepts and Their Applications in GIS

Yasin Mohammadi

GIS & AI

November 26, 2025

Presentation Roadmap

1. Data Mining
2. Modeling (General)
3. Modeling in GIS
4. Machine Learning (ML)
5. ML Algorithms
6. Classification
7. Clustering
8. Deep Learning
9. Reinforcement Learning
10. Metaheuristic Algorithms
11. Cellular Automata
12. Agents & Agent-Based Modeling

Big Picture: How These Concepts Relate

- **Data Mining**: overall process of discovering patterns in data.
- **Modeling**: representing real-world systems to explain/predict/simulate/optimize.
- **ML**: subset of AI that learns from data.
- **Deep Learning**: subset of ML using multi-layer neural networks.
- **Classification / Clustering / RL**: different ML task types.
- **Metaheuristics**: optimization search methods (not ML).
- **CA / ABM**: spatial simulation paradigms often used with GIS.

GIS perspective: All these tools convert spatial data into knowledge or decisions.

1. Data Mining

Definition: A workflow to extract patterns and knowledge from large datasets.

Step-by-step:

1. Collect spatial/non-spatial data
2. Clean and preprocess
3. Explore/visualize (EDA)
4. Apply pattern-finding methods (ML, stats, rules)
5. Interpret and deploy results

Compare:

- Broader than ML (ML is one tool inside data mining).

Example algorithm: Apriori (Association Rules)
GIS application: Mining land-use change rules from historical maps.

2. Modeling (General)

Definition: A simplified representation of reality used to describe, explain, predict, simulate, or optimize.

Key idea: Not all models are AI/ML.

Example: Linear regression model:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

GIS application: Air pollution modeling (PM2.5 prediction from traffic, weather, elevation, land cover).

3. Modeling in GIS

Definition: Modeling that explicitly includes spatial structure:

- location, distance, topology
- neighborhood effects
- spatial constraints

Compare:

- General modeling may ignore space.
- GIS modeling includes spatial dependence and rules.

Example model: Gravity / spatial interaction model:

$$F_{ij} = G\frac{P_i P_j}{d_{ij}^{\alpha}}$$

GIS application: Predicting commuting or migration flows.

4. Machine Learning (ML)

Definition: AI method where systems learn patterns from data without explicit programming.

ML workflow:

1. Define task
2. Prepare features/labels
3. Choose algorithm
4. Train model
5. Validate (accuracy metrics)
6. Deploy

Compare:

- ML $\subset$ AI
- DL $\subset$ ML

Example: Random Forest learns from labeled samples.
GIS application: Flood susceptibility mapping.

5. ML Algorithms

Definition: Specific computational methods used to learn patterns.

Families:

- **Supervised**: labeled data
- **Unsupervised**: no labels
- **Reinforcement**: reward-driven learning

Compare:

- ML is the field; algorithms are the tools.

Example algorithm: Support Vector Machine (SVM)
GIS application: Land cover classification from satellite imagery.

6. Classification (Supervised ML Task)

Definition: Predicting discrete categories (classes) from labeled data.

Steps:

1. Collect labeled samples
2. Feature extraction
3. Train classifier
4. Map prediction
5. Accuracy assessment (confusion matrix)

Compare:

- Needs labels (unlike clustering).

Example algorithm: Random Forest classifier
GIS application: Urban/vegetation/water classification from Sentinel-2.

7. Clustering (Unsupervised ML Task)

Definition: Grouping data into similar clusters without labels.

Steps:

1. Choose similarity metric
2. Choose number of clusters (if needed)
3. Run clustering
4. Interpret spatial meaning

Compare:

- No labels required.
- Often used for exploration before classification.

Example algorithm: K-means clustering
GIS application: Delineating climate zones from raster variables.

8. Deep Learning (DL)

Definition: ML using multi-layer neural networks that learn features automatically.

Why useful in GIS:

- handles imagery, LiDAR, time series
- captures complex non-linear patterns

Compare:

- DL $\subset$ ML
- Needs more data/compute but excels on raw imagery.

Example algorithm: CNN (Convolutional Neural Network)
GIS application: Building footprint extraction from aerial images.

9. Reinforcement Learning (RL)

Definition: Learning optimal actions by interacting with an environment to maximize reward.

Core parts:

- Agent, Environment, Actions, Rewards, Policy

Compare:

- RL is for sequential decision-making, not static labels.

Example algorithm: Q-learning
GIS application: Adaptive traffic signal control to reduce congestion.

10. Metaheuristic Algorithms

Definition: Optimization algorithms that search for near-optimal solutions in complex spaces.

Typical pattern:

1. Initialize random solutions
2. Iteratively improve
3. Use randomness to escape local minima

Compare:

- Not ML; does not learn from data.
- Optimizes an objective function.

Example algorithm: Genetic Algorithm (GA)
GIS application: Optimal location-allocation for fire stations.

11. Cellular Automata (CA)

Definition: Grid-based simulation where each cell updates using local neighborhood rules.

Steps:

1. Define grid and states
2. Choose neighborhood (Moore/Von Neumann)
3. Define transition rules
4. Simulate over time
5. Validate with real change

Compare:

- CA is simulation/modeling, not ML by itself.

Example: Urban Growth CA
GIS application: Simulating city expansion under slope/road/zoning constraints.

12. Agents & Agent-Based Modeling (ABM)

Definition: Simulation of autonomous agents interacting in space.

Steps:

1. Define agents + attributes
2. Define behaviors/rules
3. Define GIS environment layers
4. Run simulation
5. Analyze emergent patterns

Compare:

- CA focuses on cells; ABM focuses on individuals.
- Agents can include ML/RL internally.

Example: Pedestrian agent simulation
GIS application: Evacuation modeling and bottleneck detection.

Key Comparisons Summary

- **Data Mining vs ML:** mining is process; ML is a tool.
- **ML vs DL:** DL is ML with deep neural nets; best for imagery.
- **Classification vs Clustering:** labeled vs unlabeled tasks.
- **RL vs Others:** sequential decision learning using reward.
- **Metaheuristics vs ML:** optimization vs pattern learning.
- **CA vs ABM:** cells with local diffusion vs individuals with behaviors.

Course Plan

Topic	Sessions	Assignments	Score
CA (Cellular Automata)	2	1	30%
Agent / ABM	3	1	40%
ANN (Neural Networks)	2	1	30%

Total score: 100%

Thank You

Questions?