

Ultimate Machine Learning Flow – Expanded with ALL Topics (2025 Edition)

I. Machine Learning Foundations

- What is ML?
- Traditional vs Modern ML
- ML vs DL vs AI
- ML Lifecycle / Pipeline:

Problem → Data → Preprocessing → Algorithm Selection → Model Training → Evaluation → Optimization → Deployment → Monitoring

II. Types of Machine Learning

#	Learning Type	Key Idea	Typical Algorithms	Real-World Examples
1	Supervised Learning	Labeled data	Linear Regression, SVM, RF, NN	Spam detection, fraud
2	Unsupervised Learning	Unlabeled data	K-Means, PCA, DBSCAN	Market segmentation
3	Reinforcement Learning	Learn from environment/rewards	Q-Learning, PPO, DDPG	Game AI, robotics
4	Semi-Supervised Learning	Small labeled + large unlabeled	Ladder Networks, Label Propagation	Medical imaging
5	Self-Supervised Learning	Learns from data structure itself	BERT, SimCLR, MAE	LLMs, GenAI pre-training
6	Online Learning	Updates in real-time with streaming data	SGD, Vowpal Wabbit	Stock markets
7	Federated Learning	Decentralized training without central data	FedAvg, FedProx	Healthcare, mobile
8	Evolutionary Learning	Inspired by nature	Genetic Algorithms, Neuroevolution	NAS, RL policy search

III. Problem Types + Algorithms

Problem Type	Common Algorithms
Regression	Linear, Ridge, Lasso, RF Regressor, XGBoost
Classification	Logistic, SVM, Decision Tree, KNN, Naive Bayes
Clustering	K-Means, DBSCAN, GMM, Agglomerative
Dimensionality Reduction	PCA, t-SNE, UMAP, Autoencoders
Sequence Prediction	RNN, LSTM, GRU, Transformers
Anomaly Detection	Isolation Forest, One-Class SVM
Recommendation	Matrix Factorization, DeepFM
Reinforcement (RL)	Q-Learning, PPO, DDPG, A3C

IV. Model Training & Optimization

◆ Preprocessing

- Imputation, Scaling, Encoding
- Outlier Removal
- Feature Selection (RFE, Mutual Info)

◆ Training & Optimization

- Train/Test/Validation Split
- Cross-Validation (K-Fold, Stratified)
- Optimization:
 - Gradient Descent, SGD, Adam
- Regularization:
 - L1, L2, Dropout, Early Stopping
- Hyperparameter Tuning:
 - GridSearchCV, RandomSearchCV, Bayesian Opt, Optuna
- AutoML:
 - Auto-sklearn, H2O.ai, Google AutoML

V. Neural Networks & Deep Learning

◇ Feedforward & MLP

- Forward & Backward Propagation
- Loss: MSE, Cross-Entropy
- Activation: ReLU, Sigmoid, Tanh, Softmax

◇ Advanced Architectures

- CNNs → Image tasks
- RNNs, LSTMs → Time series/text
- Transformers → LLMs, BERT, GPT
- GANs, VAEs → Generative tasks

VI. Generative AI (GenAI)

Component	Examples
Generative Models	GANs, Diffusion Models, VAEs, LLMs
Pretraining Tasks	Masked LM (BERT), Autoregressive (GPT)
Applications	Art, Healthcare, Code Generation, Chatbots
Foundation Models	PaLM, GPT-4, Claude, Gemini
Architectures	Transformer encoder, decoder, encoder-decoder
Agents	AutoGPT, LangChain Agents, ReAct pattern

VII. Neural Architecture Search (NAS) & Hardware-Aware AI

- Manually vs Auto-designed networks

- NAS with RL or Evolutionary Algorithms
- Hardware constraints: FPGA, ASIC, GPU memory
- Co-design (model + deployment-aware optimization)

VIII. Model Evaluation Metrics

Task	Metrics
Regression	RMSE, MAE, R^2
Classification	Accuracy, Precision, Recall, F1, AUC
Clustering	Silhouette Score, DB Index
Ranking/Reco	MAP@K, NDCG, Hit Rate
RL	Cumulative Reward, Convergence

IX. Model Deployment & Monitoring

- Export: pickle, joblib, ONNX
- APIs: Flask, FastAPI, Streamlit
- Docker, Kubernetes, CI/CD
- Cloud: AWS SageMaker, Azure ML, Google Vertex AI
- Monitoring: Drift Detection, A/B Testing

X. Visualization, Tools, and Documentation

- Tools: Scikit-learn, TensorFlow, PyTorch, Hugging Face, LangChain, Gradio
- Visualization: SHAP, LIME, Confusion Matrix, Feature Importance
- Documentation: MLFlow, Weights & Biases, TensorBoard