⚙️ 12-DAY RIGOROUS MACHINE LEARNING INTERVIEW REVISION PLAN

We'll go from foundations → algorithms → optimization → deployment → interviews.

📅 Day 1 – ML Foundations & Workflow

🎯 **Goal:** Build strong intuition of how ML works end-to-end.
✅ Topics:

- ML types: Supervised, Unsupervised, Semi-supervised, Reinforcement
- ML workflow: Data → Preprocessing → Model → Evaluation → Deployment
- Bias-Variance tradeoff
- Underfitting vs Overfitting
- Train/Validation/Test split
- Cross-validation (k-fold, stratified)
- Evaluation metrics (Accuracy, F1, AUC, RMSE, MAE, R^2)

🧩 **Task:**

- Revise sklearn ML pipeline example.
- Write a quick EDA + train/test split + baseline model notebook.

📅 Day 2 – Linear & Logistic Regression

🎯 **Goal:** Revise core mathematical models deeply.
✅ Topics:

- Linear Regression: Assumptions, OLS, Gradient Descent
- Cost functions (MSE, MAE)
- Multicollinearity & VIF

- Regularization: Lasso, Ridge, ElasticNet
- Logistic Regression: Sigmoid, Log Loss, Odds ratio
- Decision boundary intuition

🧩 **Task:**

- Implement Linear & Logistic Regression *from scratch* using NumPy.
- Practice interview questions on Lasso vs Ridge intuition.

📅 Day 3 – Decision Trees & Ensemble Models (Bagging)

🎯 **Goal:** Master tree-based learners.
✅ Topics:

- Decision Trees: Entropy, Gini, Information Gain
- Overfitting in Trees + Pruning
- Random Forest: Bootstrapping, feature importance
- Bagging vs Bootstrap aggregation
- Out-of-bag error

🧩 **Task:**

- Train Decision Tree & Random Forest on Titanic dataset.
- Explain feature importance during a mock interview.

📅 Day 4 – Boosting Algorithms

🎯 **Goal:** Get complete grip on boosting family.
✅ Topics:

- Gradient Boosting (intuition + math)

- AdaBoost (weak learners, weight updates)
- XGBoost, LightGBM, CatBoost
- Hyperparameter tuning
- Difference between bagging & boosting

🧩 **Task:**

- Train & tune XGBoost on a dataset (GridSearchCV).
- Write one paragraph each explaining why boosting often wins Kaggle competitions.

📅 Day 5 – Support Vector Machines (SVM)

🎯 **Goal:** Understand the geometry & math of SVM.
✅ Topics:

- Maximum margin classifier
- Hard vs soft margin
- Kernel tricks (linear, polynomial, RBF)
- Regularization parameter (C)
- γ (gamma) effect on RBF kernel

🧩 **Task:**

- Visualize SVM decision boundaries using sklearn's `make_moons()` dataset.
- Write 5 interview Q&As on SVM hyperparameters.

📅 Day 6 – Naive Bayes, KNN, and Linear Discriminant Analysis

🎯 **Goal:** Cover remaining supervised ML fundamentals.
✅ Topics:

- Naive Bayes (Gaussian, Multinomial, Bernoulli)
- Bayes theorem intuition
- KNN algorithm (distance metrics, K selection)
- Curse of dimensionality
- LDA vs QDA (intuition + equations)

🧩 **Task:**

- Implement KNN manually (no sklearn).
- Revise one case where Naive Bayes outperforms other models.

📅 Day 7 – Unsupervised Learning (Clustering + Dimensionality Reduction)

🎯 **Goal:** Get mastery over unsupervised algorithms.
✅ Topics:

- K-Means (inertia, elbow method, silhouette score)
- Hierarchical clustering (linkage methods)
- DBSCAN
- PCA (math + explained variance)
- t-SNE, UMAP (intuition only)

🧩 **Task:**

- Perform PCA + K-Means on Iris dataset and visualize clusters.
- Explain PCA math during self-mock.

📅 Day 8 – Feature Engineering & Selection

🎯 **Goal:** Deep dive into preprocessing mastery.
✅ Topics:

- Handling missing data, outliers
- Encoding categorical variables (One-Hot, Label)
- Feature scaling (StandardScaler, MinMaxScaler)
- Feature selection: correlation, mutual info, RFE, VIF
- Dimensionality reduction overview

🧩 **Task:**

- Create a preprocessing pipeline for mixed-type data.
- Prepare answers for "How do you handle categorical data in ML?"

📅 Day 9 – Model Evaluation, Validation & Tuning

🎯 **Goal:** Learn to evaluate models like an expert.
✅ Topics:

- Confusion matrix, Precision, Recall, F1
- ROC, AUC, PR curve
- Bias-variance decomposition
- Cross-validation, GridSearchCV, RandomizedSearchCV
- Overfitting control (early stopping, dropout, regularization)

🧩 **Task:**

- Compare model performance using cross-validation.
- Write notes on when to use which metric.

📅 Day 10 – Advanced Topics & Model Interpretability

🎯 **Goal:** Prepare for senior-level questions.
✅ Topics:

- Ensemble stacking & blending
- Feature importance interpretation (SHAP, LIME)
- Model drift & monitoring
- Concept of explainable AI
- Imbalanced data handling (SMOTE, class weights)

🧩 **Task:**

- Apply SHAP or LIME on one trained model.
- Prepare an answer for "How would you handle imbalanced data?"

📅 Day 11 – Real-World ML System Design

🎯 **Goal:** Interview-style ML pipeline design.
✅ Topics:

- Problem framing (business → ML objective)
- Data collection & preprocessing pipeline
- Model selection justification
- Deployment pipeline overview (API, Docker, CI/CD)
- ML in production: monitoring & retraining

🧩 **Task:**

- Design a pipeline for fraud detection / recommendation system.
- Write 1-page "ML System Design" summary.

📅 Day 12 – Final Revision & Mock Interviews

🎯 **Goal:** Consolidate and simulate interview.
✅ Topics:

- Revise all algorithm formulas & pros/cons
- Key intuition-based comparisons (LR vs DT, RF vs XGB, etc.)
- ML puzzles and case-based questions

🧩 **Task:**

- Take one full **mock technical ML interview** (record yourself).
- Flash revise formulas + theory using handwritten notes or Anki cards.

⚡ Bonus Interview Resources

- 📘 **Book:** *Hands-On ML with Scikit-Learn & TensorFlow* – Aurélien Géron
- 🎥 **YouTube:** StatQuest (for math), Krish Naik (for implementation), CampusX (for revision)
- 🧩 **Practice:** Kaggle datasets, StrataScratch ML case questions
- 🧠 **Mock Prep:** “Machine Learning Interview Book” by Chip Huyen