✨

# 200 Days Blog Posting Challenge

> **Author:** Wonha Shin
> **Period:** *May 2024 → Sep 2025 (16 Months)*
> **Total:** 208 Days
> **Focus:** *STATistics → Machine Learning → Deep Learning → MLOps → Integration*

---

## 📘 Overview by Phase

| Phase | Days | Theme | Focus | Key Outcome |
|---|---|---|---|---|
| 🧱 **Phase I** | 1–25 | Foundations | Math, Probability, Inference | Built statistical and linear algebra foundation |
| 🤖 **Phase II** | 26–63 | Core ML | Classification, Ensemble Methods | Developed core ML modeling skills |
| 🧠 **Phase III** | 64–108 | Deep Learning | Transformers, Model Calibration | Connected DL theory with applied methods |
| 📊 **Phase IV** | 109–154 | Applied Statistics | Regression, Unsupervised Learning | Revisited classical statistical ML |
| ⚙️ **Phase V** | 155–200 | MLOps Systems | Data, Deployment, Monitoring | Designed scalable ML pipelines |
| 💡 **Phase VI** | 201–208 | Integration & Review | Coding Practice, DL Review | Consolidated skills across domains |

## 🎯 Summary Dashboard

| Category | Focus | Keywords |
|---|---|---|
| 🟦 Statistics | Probability, Regression, Hypothesis Testing | CLT, Confidence Interval, ANOVA, Regression Diagnostics, Correlation |
| 🟩 Machine Learning | Model Evaluation, Feature Engineering | PCA, LDA, SVM, Random Forest, XGBoost, Cross-Validation |
| 🟨 Deep Learning | Transformers, Fine-Tuning, Evaluation | Self-Attention, LoRA, LLaMA, Conformal Prediction, OOD Detection |
| 🟥 MLOps | System Design, Deployment, CI/CD | Dataflow, Drift Detection, Dockerization, Azure Integration, Monitoring |

# 🧩 Detailed Log (2024 – 2025)

## 📘 2024 — Building the Foundations of ML & DL

- ***From theory to intuition**: Focused on strengthening the mathematical and algorithmic backbone of machine learning.*
- *Revisited to connect statistical reasoning with core ML models and explored the foundations of deep learning architectures.*

### 🧱 Phase I – Foundations (Days 1–25)

> ***Goal:** Strengthen mathematical intuition and statistical fundamentals*

**Key Topics:**

- Linear Algebra (Matrix operations, Eigen decomposition)

- Probability Distributions (Joint, Marginal, Conditional)
- Hypothesis Testing (Z, t, Chi-square, Confidence Intervals)

| Day | Title / Topic | Date | Category | Key Concepts |
|---|---|---|---|---|
| 1 | Basic Mathematics Review (1) | May 1 2024 | 🟦 STAT | Scalar, Norm, Matrix (Inverse & Rank) |
| 2 | Basic Mathematics Review (2) | May 15 2024 | 🟦 STAT | Norm (2), Eigenvectors & Eigenvalues |
| 3 | Basic Mathematics Review (3) | May 16 2024 | 🟦 STAT | Eigendecomposition & Symmetric Matrix |
| 4 | Basic Mathematics Review (4) | May 17 2024 | 🟦 STAT | Bayesian Statistics & Var-Cov Matrix |
| 5 | Principal Component Analysis (1) | May 20 2024 | 🟩 ML | PCA Definition & Algorithm Steps |
| 6 | Principal Component Analysis (2) | May 21 2024 | 🟩 ML | PCA Applications in ML & Further Explanation |
| 7 | Linear Regression (1) | May 22 2024 | 🟩 ML | Linear Regression Concept & Equation |
| 8 | Linear Regression (2) | May 23 2024 | 🟩 ML | Cost Function & Mean Squared Error |
| 9 | Linear Regression (3) | May 24 2024 | 🟩 ML | SSE & Model Evaluation |
| 10 | Gradient in Machine Learning | May 25 2024 | 🟩 ML | Gradient Concept & Optimization Basics |
| 11 | Gradient Descent (1) | May 28 2024 | 🟩 ML | Basic Steps & Learning Rate |
| 12 | Gradient Descent (2) | May 31 2024 | 🟩 ML | Mathematical Derivation of GD |
| 13 | Gradient Descent (3) | Jun 2 2024 | 🟩 ML | Types of GD (Batch, Stochastic, Mini-Batch) |
| 14 | R-Squared (1) | Jun 3 2024 | 🟩 ML | $R^2$ Concept & Formula |
| 15 | R-Squared (2) | Jun 4 2024 | 🟩 ML | Interpreting $R^2$ as Performance Metric |
| 16 | R-Squared (3) | Jun 5 2024 | 🟩 ML | Advanced Analysis on $R^2$ |
| 17 | Regularization in ML | Jun 6 2024 | 🟩 ML | L1 & L2 Regularization, Bias-Variance Tradeoff |
| 18 | Principal Component Analysis (3) | Jun 8 2024 | 🟩 ML | Applying PCA & Scree Plot |
| 19 | Probability Review (1) | Jun 10 2024 | 🟦 STAT | Counting Principles & Random Variables |
| 20 | Probability Review (2) | Jun 11 2024 | 🟦 STAT | Joint, Marginal & Conditional Distributions |
| 21 | Probability Review (3) | Jun 14 2024 | 🟦 STAT | Continuous Distributions & Markov Chains |
| 22 | Statistics Review (1) | Jun 17 2024 | 🟦 STAT | Random Variable Properties |
| 23 | Statistics Review (2) | Jun 18 2024 | 🟦 STAT | LLN, CLT & Intro to Hypothesis Testing |
| 24 | Statistics Review (3) | Jun 19 2024 | 🟦 STAT | Z-Test, t-Test, Chi-Square Test |
| 25 | Statistics Review (4) | Jun 20 2024 | 🟦 STAT | p-Values, Confidence Intervals, Type I & II Errors |

## 🤖 Phase II – Core Machine Learning (Days 26–63)

> *Goal: Build strong ML foundations through algorithm intuition and practical implementation*

**Key Topics:**

- Linear & Logistic Regression
- Perceptron, SVM, Decision Tree, Random Forest, XGBoost
- Cross Validation, Grid Search, F1 Score, AUC

- Bias-Variance Tradeoff & Model Selection

| Day | Title / Topic | Date | Category | Key Concepts |
|---|---|---|---|---|
| 26 | Perceptron (1) | Jul 16 2024 | 🟩 ML | Single-Layer Perceptron & Training Process |
| 27 | Perceptron (2) | Jul 17 2024 | 🟩 ML | Training Steps & Decision Boundaries |
| 28 | Perceptron (3) | Jul 18 2024 | 🟩 ML | Convergence and Weight Update |
| 29 | Logistic Regression (1) | Jul 19 2024 | 🟩 ML | Sigmoid Function & Probabilistic Interpretation |
| 30 | Logistic Regression (2) | Jul 20 2024 | 🟩 ML | Cost Function & Cross-Entropy Loss |
| 31 | Logistic Regression (3) | Jul 22 2024 | 🟩 ML | Training in Scikit-Learn & Regularization |
| 32 | Support Vector Machine (1) | Jul 23 2024 | 🟩 ML | Hyperplanes & Mathematical Formulation |
| 33 | Support Vector Machine (2) | Jul 24 2024 | 🟩 ML | Non-Linear Cases & Kernel Concepts |
| 34 | Support Vector Machine (3) | Jul 25 2024 | 🟩 ML | Kernel SVM & Decision Boundaries |
| 35 | Decision Tree (1) | Jul 28 2024 | 🟩 ML | Gini Impurity & Information Gain |
| 36 | Decision Tree (2) | Jul 29 2024 | 🟩 ML | Entropy & Classification Error |
| 37 | Decision Tree (3) + Random Forest (1) | Jul 30 2024 | 🟩 ML | Tree Building & Ensemble Concepts |
| 38 | Random Forest (2) | Jul 31 2024 | 🟩 ML | Implementation & Bagging Principle |
| 39 | Linear Algebra & Matrix Review (KR) | Jul 31 2024 | 🟦 STAT | Matrix Operations & Vector Space |
| 40 | Statistics & Probability Lecture Review | Aug 1 2024 | 🟦 STAT | Comprehensive Statistics & Probability Summary |
| 41 | K-Nearest Neighbors (1) | Aug 1 2024 | 🟩 ML | Basic Concepts & Parametric vs Non-Parametric Models |
| 42 | K-Nearest Neighbors (2) | Aug 2 2024 | 🟩 ML | Distance Metrics (Euclidean, Manhattan, Minkowski, Cosine) |
| 43 | K-Nearest Neighbors (3) | Aug 5 2024 | 🟩 ML | Curse of Dimensionality |
| 44 | Data Preprocessing (1) | Aug 6 2024 | 🟩 ML | Handling Missing Data – Elimination & Imputation |
| 45 | Data Preprocessing (2) | Aug 7 2024 | 🟩 ML | Handling Categorical Data – Ordinal & One-Hot Encoding |
| 46 | Data Preprocessing (3) | Aug 8 2024 | 🟩 ML | Train/Test Split, Feature Scaling & Selection |
| 47 | Dimensionality Reduction (1) | Aug 12 2024 | 🟩 ML | PCA Overview & Linear Compression |
| 48 | Dimensionality Reduction (2) | Aug 13 2024 | 🟩 ML | Linear Discriminant Analysis (LDA) |
| 49 | Dimensionality Reduction (3) | Aug 14 2024 | 🟩 ML | Kernel PCA – Nonlinear Mapping |
| 50 | Dimensionality Reduction (4) | Aug 15 2024 | 🟩 ML | Implementing Kernel PCA in Python |
| 51 | Dimensionality Reduction (5) | Aug 17 2024 | 🟩 ML | Applying Kernel PCA to New Data |
| 52 | Pipeline in ML | Aug 18 2024 | 🟩 ML | Scikit-Learn Pipeline Concept & Example |
| 53 | Cross Validation (1) | Aug 19 2024 | 🟩 ML | Model Selection & K-Fold Cross Validation |
| 54 | Cross Validation (2) | Aug 20 2024 | 🟩 ML | Bias-Variance Tradeoff & Learning Curves |
| 55 | Cross Validation (3) | Aug 21 2024 | 🟩 ML | Grid Search & Hyperparameter Tuning |
| 56 | Cross Validation (4) | Aug 22 2024 | 🟩 ML | Confusion Matrix & F1 Score |
| 57 | Cross Validation (5) | Aug 23 2024 | 🟩 ML | ROC Curve & AUC Evaluation |

| Day | Title / Topic | Date | Category | Key Concepts |
|---|---|---|---|---|
| 58 | Class Imbalance Handling | Aug 25 2024 | 🟩 ML | Class Weights, Oversampling & Alternative Metrics |
| 59 | Ensemble Methods (1) | Aug 26 2024 | 🟩 ML | Bagging & Boosting Concepts |
| 60 | Ensemble Methods (2) | Aug 27 2024 | 🟩 ML | Majority Voting & Combining Classifiers |
| 61 | Ensemble Methods (3) | Aug 28 2024 | 🟩 ML | Majority Voting & Ensemble Evaluation |
| 62 | Ensemble Methods (4) | Aug 29 2024 | 🟩 ML | Bagging & Boosting Theory & Code |
| 63 | Ensemble Methods (5) | Aug 30 2024 | 🟩 ML | Random Forest & XGBoost Review |

## 🧠 Phase III – Deep Learning (Days 64–108)

> *Goal: Understand neural architectures and large language models; bridge theory with reliability*

**Key Topics:**

- Neural Networks (MLP, CNN, RNN, Transformer)
- Attention Mechanisms & Positional Encoding
- LoRA, Prompt Engineering, KV Cache
- Calibration, OOD Detection, Conformal Prediction

| Day | Title / Topic | Date | Category | Key Concepts |
|---|---|---|---|---|
| 64 | Deep Learning Lecture Review (1) | Sep 3 2024 | 🟨 DL | Basic Math, Supervised ML, MLP Overview |
| 65 | Deep Learning Lecture Review (2–3) | Sep 4 2024 | 🟨 DL | Neural Net Zoo: MLPs, CNNs, RNNs |
| 66 | Deep Learning Lecture Review (4) | Sep 5 2024 | 🟨 DL | CNNs, RNNs, GNNs, Inductive Bias |
| 67 | Deep Learning Lecture Review (5) | Sep 10 2024 | 🟨 DL | Transformers: GELU, LayerNorm, Architecture |
| 68 | Deep Learning Lecture Review (6) | Sep 11 2024 | 🟨 DL | BERT, GPT-2/3/4 & Transfer Learning |
| 69 | DL Review – CNNs | Sep 6 2024 | 🟨 DL | CNN Architecture & Feature Maps |
| 70 | DL Review – Transformer | Sep 7 2024 | 🟨 DL | Self-Attention, QKV, Positional Encoding |
| 71 | DL Review – NLP Overview | Sep 8 2024 | 🟨 DL | NLP Tasks, Goals, Tokenization |
| 72 | DL Lecture Review (7) | Sep 10 2024 | 🟨 DL | Algebra Refresher & ML Math Basics |
| 73 | DL Lecture Review (8) | Sep 11 2024 | 🟨 DL | Foundation Models & Transformer Workflow |
| 74 | DL Lecture Review (9) | Sep 12 2024 | 🟨 DL | Label Noise, Data Leakage, Bias Detection |
| 75–84 | Intro to NLP using Deep Learning | Oct 6 2024 | 🟨 DL | Word Embeddings, CNNs, RNNs, LSTMs, Transformer Overview |
| 85 | DL Lecture Review – Lecture 7 | Oct 7 2024 | 🟨 DL | Perplexity, Tokenizers, Embedding Layer |
| 86 | DL Lecture Review – Lecture 8 (1) | Oct 8 2024 | 🟨 DL | Text Generation, Positional Encoding, LoRA |
| 87 | DL Lecture Review – Lecture 8 (2) | Oct 9 2024 | 🟨 DL | GQA, KV Cache, MoE, DPO |
| 88 | DL Lecture Review – Lecture 10–12 | Oct 10 2024 | 🟨 DL | Numerical Precision, Hardware & Distributed Training |
| 89 | ML Review – Ensemble Method Recap | Oct 11 2024 | 🟩 ML | Random Forest & XGBoost Revisited |

| Day | Title / Topic | Date | Category | Key Concepts |
|---|---|---|---|---|
| 90 | DL Lecture Review – Background Knowledge | Oct 12 2024 | 🟨 DL | Softmax Properties, PyTorch Lightning, DataLoader |
| 91 | Optimizing Hyperparameters (W&B Integration) | Oct 14 2024 | 🟨 DL | W&B for Monitoring and Fine-Tuning ResNet-18, Post-Training Evaluations (Dying ReLU, Brightness Robustness) |
| 92 | Fine-Tuning Models (1) | Oct 17 2024 | 🟨 DL | LoRA Understanding, ResNet + SBERT for VQA |
| 93 | Fine-Tuning Models (2) & Prompt Engineering | Oct 23 2024 | 🟨 DL | Embedding Comparison (CLIP vs ResNet + SBERT), Few-Shot Prompt Design |
| 94 | Deep Learning Lecture Review – Lecture 13 | Oct 30 2024 | 🟨 DL | LLaMA 3 Architecture (RMSNorm, GQA, RoPE, SwiGLU), Pre-training & Post-training |
| 95 | Deep Learning Lecture Review – Lecture 14 | Oct 31 2024 | 🟨 DL | AI Ethics, AI Safety, AGI Challenges |
| 96 | Deep Learning Lecture Review – Lecture 15 (1) | Nov 1 2024 | 🟨 DL | Bias Detection & Mitigation, McNemar's Test, Dataset Bias |
| 97 | Deep Learning Lecture Review – Lecture 15 (2) | Nov 2 2024 | 🟨 DL | Bias Mitigation via Loss Reweighting, Sampling, Adversarial Training, DANN |
| 98 | DL Review – Revisiting LLM Models (Tree Map) | Nov 3 2024 | 🟨 DL | Transfer Learning, NLP Components, Attention Mechanism |
| 99 | Deep Learning Lecture Review – Lecture 16 | Nov 4 2024 | 🟨 DL | Uncertainty Estimation, Distribution Shift, Calibration, OOD Detection |
| 100 | Deep Learning Lecture Review – Lecture 17 (1) | Nov 12 2024 | 🟨 DL | Conformal Prediction – Marginal Coverage & Recipes |
| 101 | Deep Learning Lecture Review – Lecture 17 (2) | Nov 23 2024 | 🟨 DL | Adaptive Conformal Prediction, Calibration Set Size Analysis |
| 102 | Deep Learning Lecture Review – Lecture 18 (1) | Nov 29 2024 | 🟨 DL | Data-Centric AI – Crowdsourcing, Annotator Quality, Scaling Laws |
| 103 | Deep Learning Lecture Review – Lecture 18 (2) | Dec 9 2024 | 🟨 DL | Active Learning, SEALS, Dataset Pruning |
| 104 | Deep Learning Lecture Review – Lecture 19 | Dec 12 2024 | 🟨 DL | Label Noise, Selection Bias, Data Leakage, Subgroup Error Analysis |
| 105 | Deep Learning Lecture Review – Lecture 20 (End) | Dec 18 2024 | 🟨 DL | Model Drift Detection, Periodic Retraining, Continual Learning |
| 106 | Enhancing DL Workflows (Data Distribution) | Dec 19 2024 | 🟨 DL | Optimization via Data Loading, Profiling, DDP vs DP |
| 107 | HW4 – Model Calibration & Conformal Prediction | Dec 20 2024 | 🟨 DL | Platt Scaling, Label Smoothing, Naïve & Adaptive Prediction Sets |
| 108 | HW5 – OOD Detection & Continual Learning | Dec 21 2024 | 🟨 DL | Maximum Softmax Probability, ODIN, SLDA, IID Streaming |

## ⚙️ 2025 — Starting with MLOps & Statistical Integration

- ***From learning models to building systems:*** *Transitioned from individual model understanding to end-to-end ML systems.*
- *Focused on automation, reliability, and data-centric workflows in production-level MLOps pipelines.*

## 📊 Phase IV – Applied Statistics (Days 109–154)

> *Goal: Connect classical statistical inference with modern ML applications*

**Key Topics:**

- Regression (OLS, Interactions, Multicollinearity)
- Classification (Naïve Bayes, LDA, Logistic Regression)
- Unsupervised Learning (PCA, K-Means, GMM, Gower's Distance)
- Model Evaluation & Interpretability

| Day | Title / Topic | Date | Category | Key Concepts |
|---|---|---|---|---|
| 109 | Exploratory Data Analysis (1) | Jan 25 2025 | 🟦 STAT | Data Types, Descriptive Statistics, Measures of Location & Variability |
| 110 | Exploratory Data Analysis (2) | Jan 26 2025 | 🟦 STAT | Correlation, Distribution, Visualization Methods |
| 111 | Data & Sampling Distributions (1) | Jan 28 2025 | 🟦 STAT | Sampling, Bias, Central Limit Theorem |
| 112 | Data & Sampling Distributions (2) | Jan 29 2025 | 🟦 STAT | Bootstrap, Confidence Intervals, Normal Distribution |
| 113 | Data & Sampling Distributions (3) | Jan 31 2025 | 🟦 STAT | t-, Chi-Square, F-, Poisson Distributions |
| 114 | Statistical Experiments & Significance Testing (1) | Feb 1 2025 | 🟦 STAT | A/B Testing, Hypothesis Tests, Permutation Tests |
| 115 | Statistical Experiments & Significance Testing (2) | Feb 5 2025 | 🟦 STAT | p-Values, Type I & II Errors, Practical Applications |
| 116 | Statistical Experiments & Significance Testing (3) | Feb 6 2025 | 🟦 STAT | t-Tests, Multiple Testing, False Discovery Rate |
| 117 | Statistical Experiments & Significance Testing (4) | Feb 7 2025 | 🟦 STAT | ANOVA (One- & Two-Way), F-Statistics, Chi-Square |
| 118 | Statistical Experiments & Significance Testing (5) | Feb 9 2025 | 🟦 STAT | Chi-Square Theory, Fisher's Exact Test, Power & Sample Size |
| 119 | Regression & Prediction (1) | Feb 11 2025 | 🟦 STAT | Simple & Multiple Linear Regression, Least Squares |
| 120 | Regression & Prediction (2) | Feb 13 2025 | 🟦 STAT | Model Selection, Cross-Validation, Prediction Using Regression |
| 121 | Regression & Prediction (3) | Feb 15 2025 | 🟦 STAT | Dummy Variables, Factor Variables, Interactions |
| 122 | Regression & Prediction (4) | Feb 16 2025 | 🟦 STAT | Multicollinearity, Confounding Variables, Interpretation |
| 123 | Regression & Prediction (5) | Feb 18 2025 | 🟦 STAT | Outliers, Influential Obs., Heteroskedasticity |
| 124 | Regression & Prediction (6) | Feb 19 2025 | 🟦 STAT | Polynomial Regression, Splines, GAM |
| 125 | Revisiting Mathematics Theories (1) | Feb 22 2025 | 🟦 STAT | Hypothesis Testing, Paired/Independent Samples |
| 126 | Revisiting Mathematics Theories (2) | Feb 23 2025 | 🟦 STAT | Chi-Square & F-Distribution, Inference for Variance |
| 127 | Revisiting Mathematics Theories (3) | Feb 24 2025 | 🟦 STAT | Proportion Inference, Chi-Square Test |
| 128 | Revisiting Mathematics Theories (4) | Feb 25 2025 | 🟦 STAT | Non-Parametric Inference (Wilcoxon Tests) |
| 129 | Regression & Prediction (7) | Feb 26 2025 | 🟦 STAT | Stepwise Regression & Model Selection |
| 130 | Regression & Prediction (8) | Feb 28 2025 | 🟦 STAT | Weighted Regression & Interaction Effects |
| 131 | Classification (1) | Mar 1 2025 | 🟦 STAT | Naïve Bayes – Theoretical Approach & Prediction |
| 132 | Classification (2) | Mar 2 2025 | 🟦 STAT | Discriminant Analysis – Covariance, Default Risk |
| 133 | Classification (3) | Mar 3 2025 | 🟦 STAT | Logistic Regression – Mathematical Foundation (Odds, Logit) |

| Day | Title / Topic | Date | Category | Key Concepts |
|---|---|---|---|---|
| 134 | Classification (4) | Mar 5 2025 | 🟦 STAT | Logistic Regression – GLM, Fitting, Interpretation |
| 135 | Classification (5) | Mar 7 2025 | 🟦 STAT | Logistic Regression – Model Assessment |
| 136 | Classification (6) | Mar 8 2025 | 🟦 STAT | Evaluating Classifiers – Confusion Matrix, ROC, Lift |
| 137 | Classification (7) | Mar 9 2025 | 🟦 STAT | Handling Imbalanced Data – Over/Undersampling |
| 138 | Classification (8) | Mar 14 2025 | 🟦 STAT | Strategies for Imbalanced Data – SMOTE, Cost-Based Classification |
| 139 | Statistical ML (1) | Mar 15 2025 | 🟦 STAT | KNN – Distance Metrics & One-Hot Encoding |
| 140 | Statistical ML (2) | Mar 16 2025 | 🟦 STAT | KNN – Choosing K, Standardization |
| 141 | Statistical ML (3) | Mar 20 2025 | 🟦 STAT | KNN as Feature Engine + Tree Model Intro |
| 142 | Tree Models (2) | Mar 21 2025 | 🟦 STAT | Recursive Partitioning, Impurity (Gini/Entropy) |
| 143 | Tree Models (3) | Mar 25 2025 | 🟦 STAT | Overfitting Handling, Continuous Prediction |
| 144 | Ensemble (1) | Mar 28 2025 | 🟦 STAT | Bagging & Random Forest – Core Idea |
| 145 | Ensemble (2) | Mar 29 2025 | 🟦 STAT | Random Forest (2) – Variable Importance |
| 146 | Boosting (1) | Mar 30 2025 | 🟦 STAT | AdaBoost & XGBoost Key Concepts |
| 147 | Boosting (2) | Mar 31 2025 | 🟦 STAT | Regularization, Hyperparameters, Cross-Validation |
| 148 | Unsupervised Learning (1) | Apr 2 2025 | 🟦 STAT | PCA – Computing Principal Components |
| 149 | Unsupervised Learning (2) | Apr 4 2025 | 🟦 STAT | PCA – Interpretation & Correspondence Analysis |
| 150 | Unsupervised Learning (3) | Apr 6 2025 | 🟦 STAT | K-Means Clustering – Algorithm & Example |
| 151 | Unsupervised Learning (4) | Apr 8 2025 | 🟦 STAT | K-Means: Determining Optimal K, Interpretation |
| 152 | Unsupervised Learning (5) | Apr 13 2025 | 🟦 STAT | Hierarchical Clustering – Dendrogram, Linkage Methods |
| 153 | Unsupervised Learning (6) | Apr 15 2025 | 🟦 STAT | Model-Based Clustering – Gaussian Mixture Models |
| 154 | Unsupervised Learning (7) | Apr 20 2025 | 🟦 STAT | Scaling, Mixed Data, Gower's Distance |

## ⚙️ Phase V – MLOps Systems (Days 155–203)

> *Goal: Learn scalable ML systems design from data engineering to deployment & monitoring*

**Key Topics:**

- Data Engineering (ETL, Parquet, Streaming Pipelines)
- Feature Engineering & Data Leakage Prevention
- Experiment Tracking & Model Versioning (W&B)
- Model Serving (Batch vs Online, Cloud vs Edge)
- Monitoring & Drift Detection (Prometheus, Grafana)
- CI/CD Automation (Docker, GitHub Actions, Azure ACI)

| Day | Title / Topic | Date | Category | Key Concepts |
|---|---|---|---|---|
| 155 | Intro to ML System Design (1) | Apr 24 2025 | 🟥 MLOps | Business ↔ ML Objectives, Requirements, Iterative Design |
| 156 | Intro to ML System Design (2) | Apr 25 2025 | 🟥 MLOps | Problem Framing, Task Definition, Evaluation Criteria |
| 157 | Intro to ML System Design (3) | Apr 26 2025 | 🟥 MLOps | ML Task Types, Objective Functions |
| 158 | Data Engineering Fundamentals (1) | Apr 29 2025 | 🟥 MLOps | Data Formats (JSON, Parquet), Storage Engines, ETL |
| 159 | Data Engineering Fundamentals (2) | May 17 2025 | 🟥 MLOps | Dataflow Modes – Batch vs Streaming |
| 160 | Training Data (1) | May 18 2025 | 🟥 MLOps | Sampling Methods – Stratified, Weighted, Reservoir |
| 161 | Training Data (2) | May 23 2025 | 🟥 MLOps | Labeling – Hand vs Natural Labels, Active Learning |
| 162 | Training Data (3) | May 25 2025 | 🟥 MLOps | Class Imbalance – Metrics, Sampling, Algorithm-Level Fixes |
| 163 | Training Data (4) | May 26 2025 | 🟥 MLOps | Data Augmentation – Label-Preserving Transformations |
| 164 | Feature Engineering (1) | May 29 2025 | 🟥 MLOps | Handling Missing Values, Scaling, Encoding, Normalization |
| 165 | Feature Engineering (2) | May 31 2025 | 🟥 MLOps | Advanced Features – Positional Embedding, Domain-Specific Engineering |
| 166 | Feature Engineering (3) | Jun 2 2025 | 🟥 MLOps | Data Leakage – Detection & Prevention |
| 167 | Model Development & Evaluation (1) | Jun 27 2025 | 🟥 MLOps | Model Selection, Ensemble (Bagging/Boosting/Stacking) |
| 168 | Model Development & Evaluation (2) | Jun 28 2025 | 🟥 MLOps | Experiment Tracking, Versioning, Distributed Training |
| 169 | Model Development & Evaluation (3) | Jun 29 2025 | 🟥 MLOps | AutoML, NAS, Establishing Baselines |
| 170 | Model Development & Evaluation (4) | Jun 30 2025 | 🟥 MLOps | Offline Evaluation – Perturbation & Invariance Tests |
| 171 | Model Deployment & Prediction Service (1) | Jul 1 2025 | 🟥 MLOps | Batch vs Online Prediction, Unified Pipeline Design |
| 172 | Model Deployment & Prediction Service (2) | Jul 2 2025 | 🟥 MLOps | Model Compression – Low-Rank Factorization, KD, Pruning, Quantization |
| 173 | Model Deployment & Prediction Service (3) | Jul 3 2025 | 🟥 MLOps | Cloud vs Edge Deployment, Model Optimization (AutoTVM, WebAssembly) |
| 174 | Data Distribution Shifts & Monitoring (1) | Jul 5 2025 | 🟥 MLOps | ML System Failures – Feedback Loops & Edge Cases |
| 175 | Data Distribution Shifts & Monitoring (2) | Jul 7 2025 | 🟥 MLOps | Covariate Shift, Label Shift, Concept Drift Detection |
| 176 | Data Distribution Shifts & Monitoring (3) | Jul 8 2025 | 🟥 MLOps | Addressing Drifts – Importance Weighting & Adaptation |
| 177 | Data Distribution Shifts & Monitoring (4) | Jul 9 2025 | 🟥 MLOps | Monitoring vs Observability, ML Metrics & System Reliability |
| 178-190 | SQL Fundamentals for Data Analysis (Course Completed) | Jul 11 2025 | 🟥 MLOps | SQL Basics, Joins, Analytic Functions (Window, Ranking) |
| 191-198 | Hands-On Docker (Course Completed) | Aug 11 2025 | 🟥 MLOps | Docker CLI → Compose → Multi-Container Orchestration |
| 199-200 | CI/CD with GitHub Actions + Azure (1) | Sep 2 2025 | 🟥 MLOps | Build & Push Docker Images → Deploy to Azure ACI via ACR Automation |
| 201–203 | CI/CD with GitHub Actions + Azure (2) | Sep 5–8 2025 | 🟥 MLOps | Full CI/CD workflow with secret management, ACI → AKS scaling |

## 💡 Phase VI – Integration & Review (Day 204–208)

> Theme: Leetcode + DL Review Integration
>
> **Focus:** Strengthening problem-solving and system intuition.

**Highlights:**

- Python & SQL Problem Solving (Two Sum, Max Subarray, Joins)
- DL Concept Review – CNNs, Transformers, Positional Encoding
- Revisiting Optimizers (SGD, Adam), Autoencoders, Knowledge Distillation
- Connecting Algorithms with Deep Learning Insights

🟦 *Core Skills*: Problem Solving • Systems Thinking • Cross-Domain Integration

| Day | Title / Topic | Date | Category | Key Concepts |
|---|---|---|---|---|
| 204 | DL Review – Optimizers, CNNs & Data Drift | Sep 9 2025 | 🟨 DL | Optimizers (SGD, Adam), Parameter Sharing in CNNs, Concept Drift Overview |
| 205 | Python 175 & SQL Inner Join & DL Review | Sep 12 2025 | 🟩 ML / 🟨 DL | Two Sum (Python), SQL Joins, Embedding Layers, Autoencoders, Knowledge Distillation |
| 206 | Python 217 & SQL 175/176 & DL Review | Sep 13 2025 | 🟩 ML / 🟨 DL | Contains Duplicate, Second-Highest Salary, Transfer Learning, Fine-Tuning, CNNs |
| 207 | Python 53 & SQL 185 & DL Review | Sep 15 2025 | 🟩 ML / 🟨 DL | Maximum Subarray, Department Top Sales, RNNs, LSTMs, Vanishing/Exploding Gradients |
| 208 | Python 121 & SQL 175/176 & DL Review | Sep 16 2025 | 🟩 ML / 🟨 DL | Best Time to Buy/Sell Stock, Second-Highest Salary, Transformers & Positional Encoding |