



The Complete Machine Learning Bootcamp (5-Month Syllabus)

Module 1: Foundations & The ML Lifecycle (Weeks 1-2)

- C1: Course Overview & The ML Landscape: AI vs ML vs DL. Types of ML (Supervised, Unsupervised, RL). Real-world Applications. Challenges.
- C2: Roles & Process: Data Engineer vs Analyst vs Scientist vs ML Engineer. The Machine Learning Development Life Cycle (MLDLC).
- C3: Tools & Environment: Installing Anaconda, Jupyter Lab, Intro to Google Colab. Python/NumPy Refresher. What are Tensors? (In-depth explanation).
- Assignment 1: Set up environment and write a report comparing the data roles.

Module 2: Data Acquisition & EDA (Weeks 2-4)

- C4: End-to-End Toy Project: How to Frame an ML Problem. Working with CSV files. Building a first model (KNN).
- C5: Data Sources - Structured: Advanced CSV handling. Working with JSON data. Reading from SQL databases.
- C6: Data Sources - Unstructured: Acquiring Data via APIs (requests library). Web Scraping with BeautifulSoup to create a DataFrame.
- Assignment 2: Build a dataset from two different sources (e.g., API + CSV).
- C7: Understanding Data - Descriptive Statistics: Using Pandas for deep profiling.
- C8: Univariate Analysis: Visualizing single variables (Histograms, Box Plots, KDE).
- C9: Bivariate Analysis: Visualizing relationships (Scatter Plots, Pairplots, Correlation Matrices).
- Assignment 3: Perform a complete statistical and visual EDA on your acquired dataset.
- C10: Multivariate Analysis: Advanced visualizations (Heatmaps, 3D plots, grouped analyses).
- C11: Automated EDA: Using pandas-profiling for fast insights.
- C12: Feature Engineering Introduction & Feature Scaling - Standardization.
- Assignment 4: Generate an automated EDA report. Apply standardization to numerical features.

Module 3: Core Feature Engineering II (Weeks 5-7)

- C13: Feature Scaling: Normalization (MinMax), MaxAbsScaling, RobustScaling.
- C14: Encoding Categorical Data: Ordinal Encoding & Label Encoding.
- C15: Encoding Categorical Data: One-Hot Encoding & Handling High Cardinality.
- Assignment 5: Preprocess a dataset: scale numerical features and encode categorical ones.
- C16: The ColumnTransformer: Applying different transformations to different columns efficiently.
- C17: Sklearn Pipelines A-Z: Building end-to-end pipelines to avoid data leakage.
- C18: Function Transformers: Log Transform, Reciprocal Transform, Square Root Transform for managing skew.
- Assignment 6: Build a pipeline that uses ColumnTransformer and a simple model.
- C19: Power Transforms: Box-Cox Transform & Yeo-Johnson Transform.
- C20: Binning & Discretization: Quantile Binning, K-Means Binning. Binarization.
- C21: Handling Mixed Variables.
- Assignment 7: Apply power transforms to fix skewness and create new binned features.

Module 4: Advanced Feature Engineering (Weeks 8-9)

- C22: Handling Date & Time Variables: Extracting features (dayofweek, hour, is_weekend, etc.).
- C23: Handling Missing Data - Theory: MCAR, MAR, MNAR. Complete Case Analysis.
- C24: Handling Missing Numerical Data: Mean/Median/Mode Imputation (SimpleImputer).
- Assignment 8: Handle missing values in your dataset and extract features from datetime columns.
- C25: Handling Missing Categorical Data: Most Frequent Imputation, Adding a "Missing" Category.
- C26: Advanced Imputation I: Missing Indicator & Random Sample Imputation.
- C27: Advanced Imputation II: Multivariate Imputation with KNNImputer.
- Assignment 9: Compare a simple imputation strategy with KNN imputation.
- C28: Advanced Imputation III: Iterative Imputation with IterativeImputer (MICE).
- C29: What are Outliers? Why they are a problem. Z-Score Method.
- C30: Outlier Handling: IQR Method. Percentile Method (Winsorization).
- Assignment 10: Detect and cap outliers using the IQR method on your dataset.



Research Innovation & Thesis Center

Dhanmondi 9/A, Dhaka-1209, Bangladesh

Module 5: Dimensionality & Linear Models (Weeks 10-12)

- C31: Feature Construction & Feature Splitting.
- C32: The Curse of Dimensionality.
- C33: Dimensionality Reduction - PCA: Geometric Intuition & Mathematical Formulation.
- Assignment 11: Apply PCA for visualization and dimensionality reduction.
- C34: PCA - Code Implementation & Visualization.
- C35: Simple Linear Regression: Intuition, Mathematical Formulation, Coding from Scratch.
- C36: Regression Metrics: MSE, MAE, RMSE, R^2 , Adjusted R^2 .
- Assignment 12: Implement Simple Linear Regression from scratch and evaluate it.
- C37: Multiple Linear Regression: Intuition & Code with sklearn.
- C38: Multiple Linear Regression: Mathematical Formulation from Scratch.
- C39: Gradient Descent from Scratch: The core algorithm behind most ML (with animation).
- Assignment 13: Implement Multiple Linear Regression and Gradient Descent from scratch.

Module 6: Advanced Regression & Model Tuning (Weeks 13-14)

- C40: Types of Gradient Descent: Batch, Stochastic, Mini-Batch.
- C41: Polynomial Regression.
- C42: Bias-Variance Trade-off: Overfitting and Underfitting.
- Assignment 14: Demonstrate overfitting with polynomial regression and discuss bias-variance.
- C43: Regularization - Ridge Regression (L_2): Intuition, Code, Maths.
- C44: Regularization - Lasso Regression (L_1): Intuition, Code, Why it creates sparsity.
- C45: Regularization - ElasticNet: Intuition and Code.
- Assignment 15: Compare Linear, Ridge, Lasso, and ElasticNet models on a dataset.



Research Innovation & Thesis Center

Dhanmondi 9/A, Dhaka-1209, Bangladesh

Module 7: Classification & Evaluation (Weeks 15-16)

- C46: Logistic Regression I: Perceptron Trick, Sigmoid Function.
- C47: Logistic Regression II: Loss Function (Log Loss/Cross-Entropy), MLE.
- C48: Logistic Regression III: Gradient Descent & Code from Scratch.
- Assignment 16: Implement Logistic Regression from scratch.
- C49: Classification Metrics I: Accuracy, Confusion Matrix, Type I/II Error.
- C50: Classification Metrics II: Precision, Recall, F1-Score.
- C51: Logistic Regression IV: Softmax/Multinomial Regression for multi-class.
- Assignment 17: Evaluate a classifier using a full suite of metrics and interpret them.
- C52: Logistic Regression V: Polynomial Features & Hyperparameter Tuning.
- C53: Introduction to Ensemble Learning: Bagging vs Boosting.
- C54: Voting Ensembles: Hard Voting vs Soft Voting (Classifier & Regressor).
- Assignment 18: Build a Voting Classifier and compare it to a base Logistic model.

Module 8: Trees, Ensembles & Unsupervised Learning (Weeks 17-18)

- C55: Decision Trees I: Geometric Intuition, Entropy, Gini, Information Gain.
- C56: Decision Trees II: Hyperparameters, Overfitting, Regression Trees.
- C57: Decision Trees III: Visualization with dtreeviz.
- Assignment 19: Build, tune, and visualize a Decision Tree.
- C58: Bagging Ensembles: Introduction, Bagging Classifier/Regressor.
- C59: Random Forest I: Intuition, Bias-Variance, Why it works so well.
- C60: Random Forest II: Bagging vs Random Forest, Hyperparameters, OOB Score.
- Assignment 20: Compare a Decision Tree to a Bagging model and a Random Forest.
- C61: Random Forest III: Hyperparameter Tuning with GridSearchCV/RandomizedSearchCV.
- C62: Random Forest IV: Feature Importance.
- C63: Boosting - AdaBoost: Intuition, Step-by-Step Explanation, Code.
- Assignment 21: Perform hyperparameter tuning on a Random Forest and interpret feature importance.
- C64: Boosting - Gradient Boosting: Intuition for Regression & Classification.
- C65: Introduction to XGBoost.
- C66: Unsupervised Learning - K-Means Clustering: Intuition, Algorithm, Code.
- Assignment 22: Implement K-Means clustering for customer segmentation.
- C67: Advanced Ensembles: Stacking and Blending.
- C68: Course Capstone & Review: Final project workshop, Q&A, next steps.
- Final Project: End-to-end project covering all steps of the MLDLC on a complex dataset.