

Data Preprocessing <ul style="list-style-type: none">• Normalization: $x = \frac{x - \min(x)}{\max(x) - \min(x)}$• Standardization: $x = \frac{x - \mu}{\sigma}$• Handling Missing Values:<ul style="list-style-type: none">– Mean/Median Imputation– Mode Imputation– KNN Imputation	Feature Engineering <ul style="list-style-type: none">• Polynomial Features: $x_1, x_2, x_1^2, x_1 x_2, x_2^2$• Interaction Features: $x_1 \cdot x_2$• Log Transform: $\log(x)$• Binning: Discretize continuous variables Model Selection <ul style="list-style-type: none">• Linear Regression: $y = \beta_0 + \beta_1 x + \epsilon$	Regression: <ul style="list-style-type: none">• Logistic Regression: $\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$• Decision Trees: Recursive binary splitting• Random Forest: Ensemble of decision trees• SVM: Maximize margin between classes Model Evaluation <ul style="list-style-type: none">• Confusion Matrix:	<ul style="list-style-type: none">– TP, FP, TN, FN• Accuracy: $\frac{TP+TN}{TP+TN+FP+FN}$• Precision: $\frac{TP}{TP+FP}$• Recall: $\frac{TP}{TP+FN}$• F1 Score: $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$• ROC Curve: Plot of TPR vs. FPR	Hyperparameter Tuning <ul style="list-style-type: none">• Grid Search: Exhaustive search over parameter grid• Random Search: Randomly sample parameters• Bayesian Optimization: Model-based optimization• Cross-Validation: k-Fold, Leave-One-Out
---	---	--	---	---