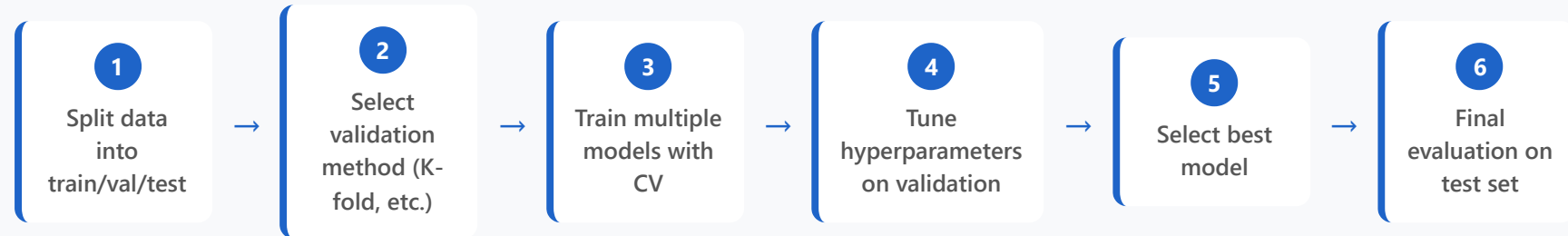


# Model Selection Strategy

## Complete Model Selection Workflow



## Recommended Data Split



Training: Model learning • Validation: Hyperparameter tuning & model selection • Test: Final unbiased evaluation



### Never Touch Test Set

Test set must remain unseen until final evaluation



### Use Cross-Validation

Apply CV on training set for robust estimates



### Consider Data Balance

Use stratified splits for imbalanced datasets



### Monitor Overfitting

Compare train vs validation performance

## 🌟 Key Principles



Choose metrics aligned with business goals



Use multiple metrics for comprehensive view



Balance performance with computational cost



Analyze errors to understand model behavior



Document all decisions and experiments



Iterate based on validation results

## 📋 Model Performance Comparison

Model	Best Hyperparameters	Accuracy	Precision	Recall	F1 Score	Training Time	Status
Random Forest	<code>n_estimators: 200</code> <code>max_depth: 15</code> <code>min_samples_split: 5</code>	0.942	0.938	0.945	0.941	2.3s	✓ Best
XGBoost	<code>learning_rate: 0.1</code> <code>max_depth: 8</code> <code>n_estimators: 150</code>	0.935	0.932	0.937	0.934	1.8s	2nd
SVM (RBF)	<code>C: 10</code> <code>gamma: 0.01</code> <code>kernel: rbf</code>	0.918	0.915	0.920	0.917	5.1s	-
Logistic Regression	<code>C: 1.0</code> <code>penalty: l2</code> <code>solver: lbfgs</code>	0.887	0.883	0.891	0.887	0.5s	⚡ Fastest
Decision Tree	<code>max_depth: 12</code> <code>min_samples_split: 10</code> <code>criterion: gini</code>	0.865	0.861	0.869	0.865	0.3s	-

💡 All metrics evaluated on validation set using 5-fold cross-validation. Best model selected based on F1 score and generalization performance.