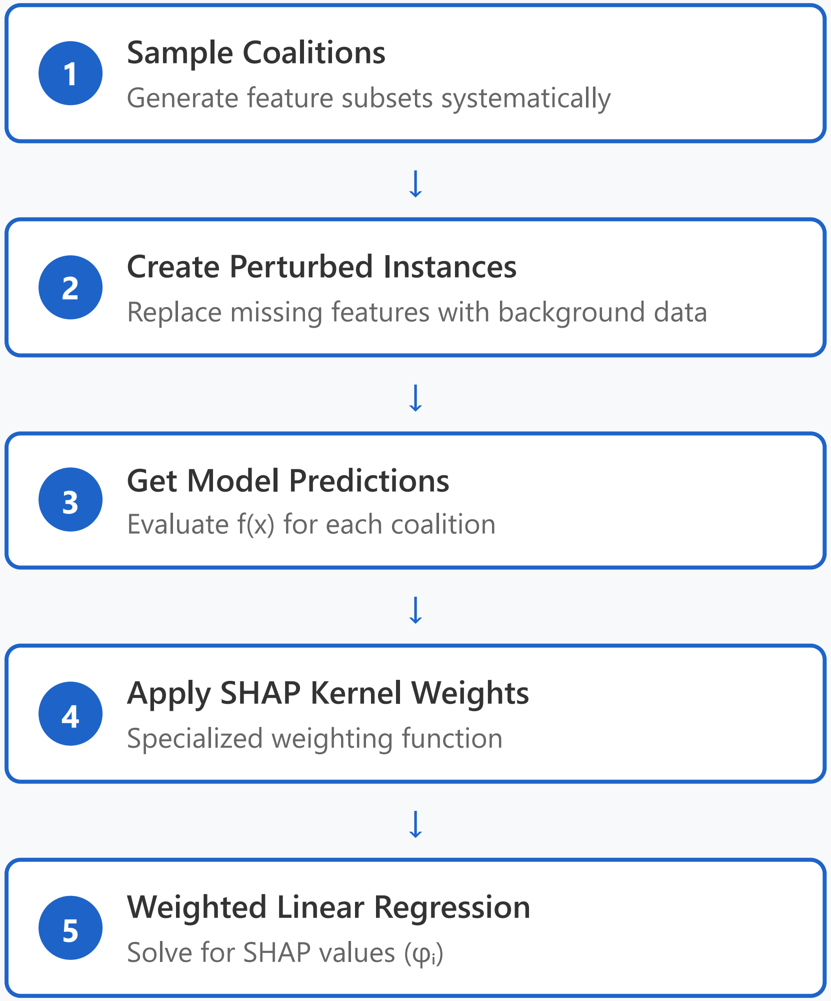
KernelSHAP: Model-Agnostic Approximation

Sampling-based method with weighted linear regression

Algorithm Flowchart



1 **Sample Coalitions**
Generate feature subsets systematically

↓

2 **Create Perturbed Instances**
Replace missing features with background data

↓

3 **Get Model Predictions**
Evaluate f(x) for each coalition

↓

4 **Apply SHAP Kernel Weights**
Specialized weighting function

↓

5 **Weighted Linear Regression**
Solve for SHAP values (φ_i)

🎯

Key Advantage

Model-Agnostic - Works with any black-box model

- Neural networks
- Ensemble methods
- Custom models

⚖️

Trade-off

Computational Cost vs Accuracy

More samples → Better approximation
Fewer samples → Faster computation

💻

Implementation

```python
import shap

explainer = shap.KernelExplainer(
    model.predict,
    X_train
)
shap_values = explainer.shap_values(X_test)
```