QuickVina Machine Learning Demo
100 Pharmaceutical Compounds Analysis

MACHINE LEARNING MOLECULAR DOCKING DEMONSTRATION

Dataset: 100 diverse pharmaceutical compounds
Chemical Space: Analgesics, CNS drugs, cardiovascular, antibiotics
Valid Binding Results: 92 compounds
Fingerprint Method: Morgan Circular Fingerprints (ECFP)
Fingerprint Dimensions: 2048 bits (radius=2)

BINDING AFFINITY STATISTICS:
Range: -10.17 to -3.39 kcal/mol
Mean: -6.39 ± 1.03 kcal/mol
Median: -6.31 kcal/mol

MACHINE LEARNING MODEL PERFORMANCE:
Algorithm: XGBoost Gradient Boosting Regressor
Features: Morgan Fingerprints (2048-dimensional binary vectors)
Training Strategy: 80/20 train/test split with 3-fold CV
Feature Engineering: StandardScaler normalization

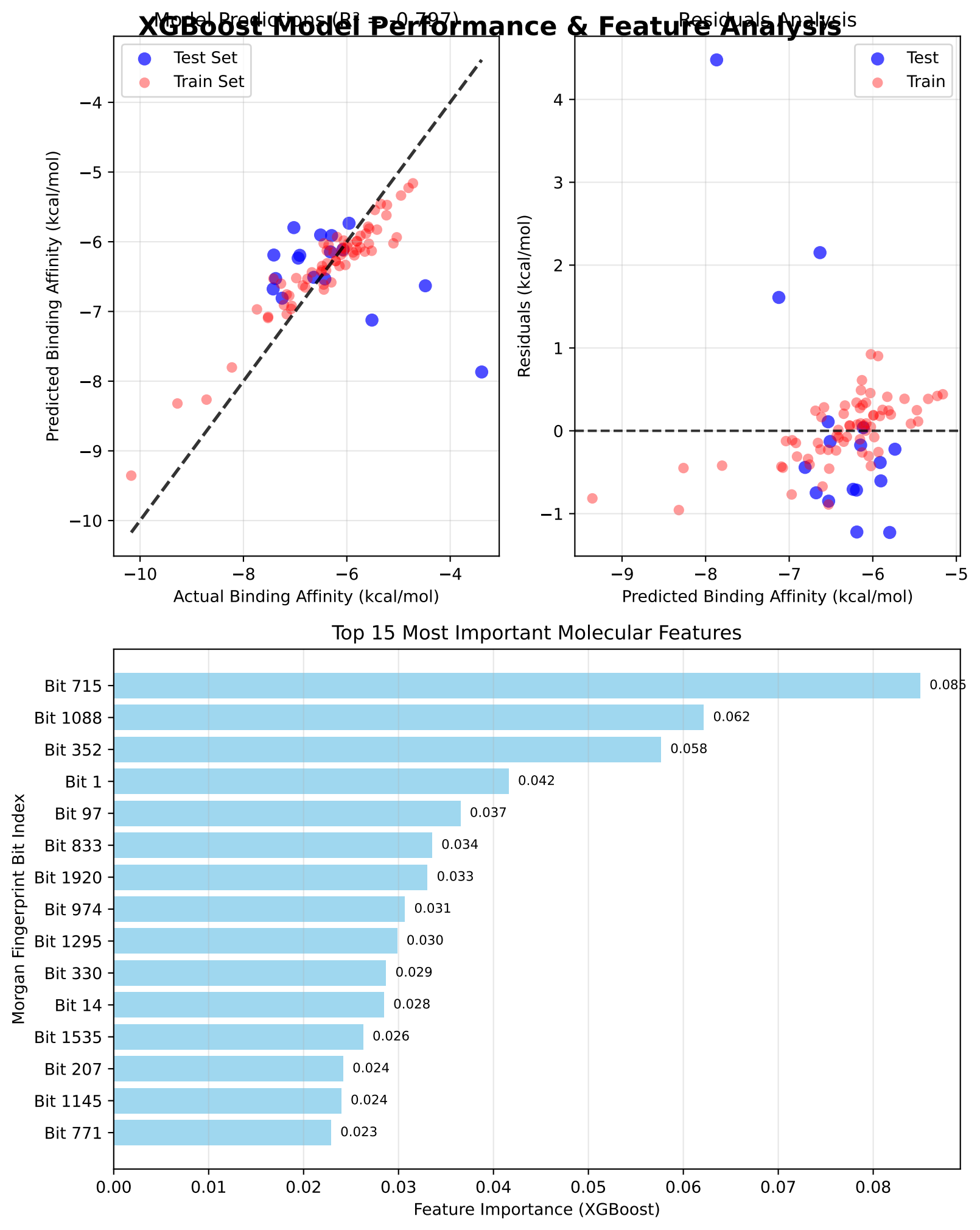
MODEL PERFORMANCE METRICS:
• R^2 Score (Test Set): -0.797
• Root Mean Square Error: 1.401 kcal/mol
• Mean Absolute Error: 0.930 kcal/mol
• Cross-Validation R^2: -1.180 ± 1.091

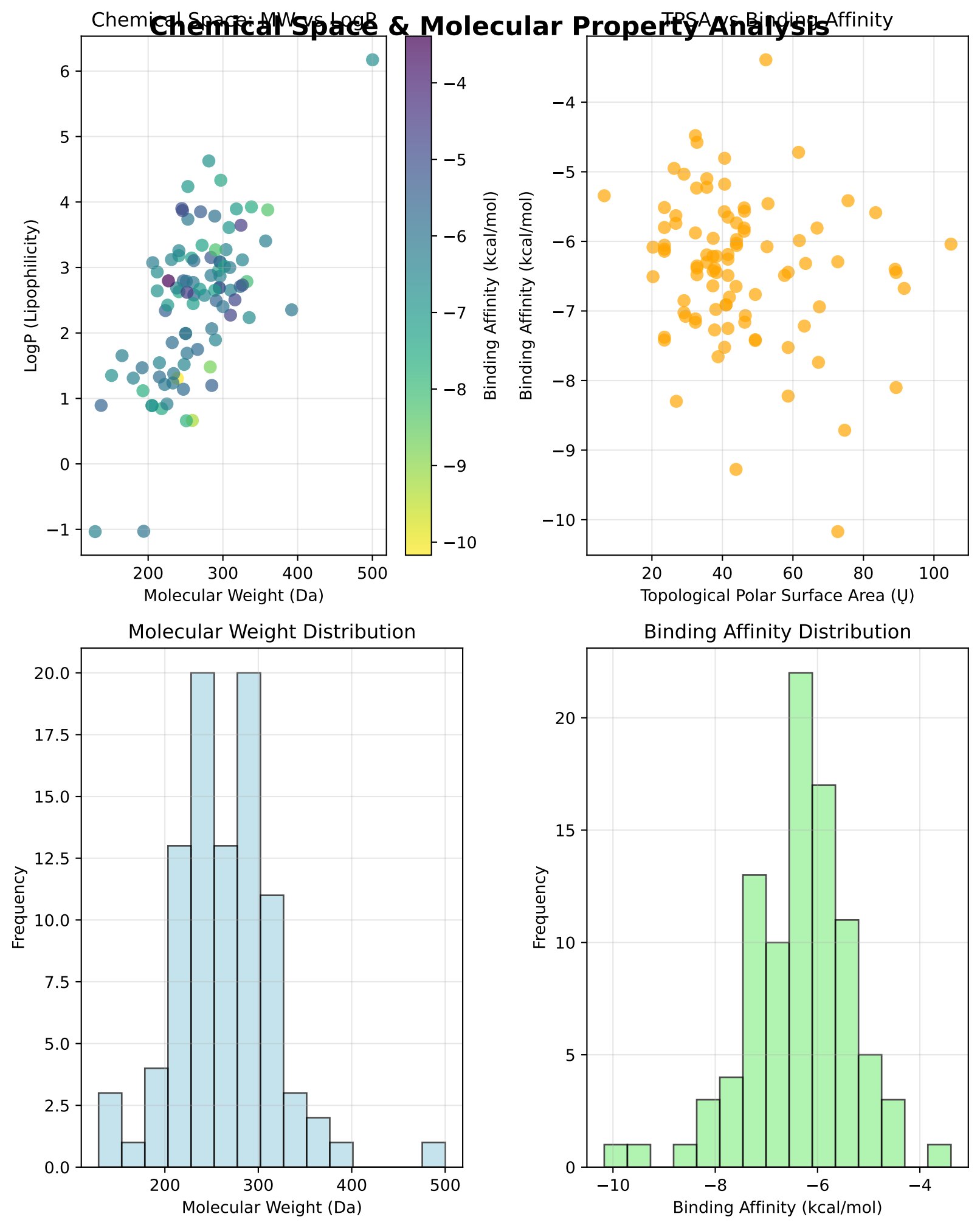
OPTIMIZED HYPERPARAMETERS:
• Number of Estimators: 100
• Maximum Tree Depth: 3
• Learning Rate: 0.1
• Subsample Ratio: 1.0

DATASET CHARACTERISTICS:
• Feature Dimensions: 2048

TOP 10 BINDING AFFINITIES:

Rank	Affinity (kcal/mol)	SMILES
1	-10.17	CC(C)(C)NCC(C1=CC(=C(C=C1...
2	-9.28	CN1C(=O)C2=C(C1=O)C=CC(=C...
3	-8.71	CC1=CC=C(C=C1)S(=O)(=O)N2...
4	-8.30	CN1C2=CC=CC=C2C(=O)N(C1=S...
5	-8.22	CC(C)(C)OC(=O)NC1=CC=C(C=...
6	-8.10	CC(C)CC1=CC=C(C=C1)C(=O)N...
7	-7.74	CN1CCN(CC1)C(=O)C2=CC=C(C...
8	-7.66	CN(C)C(=O)C1=CC2=C(C=C1)O...
9	-7.53	CC(C)(C)OC(=O)N1CCC(CC1)C...
10	-7.52	CN(C)CCN1C(=O)C2=CC=CC=C2...


XGBoost Model Performance & Feature Analysis
Model Predictions (R^2 = 0.797)
Test Set
Train Set
Predicted Binding Affinity (kcal/mol)
Actual Binding Affinity (kcal/mol)
Residuals Analysis
Test
Train
Residuals (kcal/mol)
Predicted Binding Affinity (kcal/mol)
Top 15 Most Important Molecular Features
Bit 715
Bit 1088
Bit 352
Bit 1
Bit 97
Bit 833
Bit 1920
Bit 974
Bit 1295
Bit 330
Bit 14
Bit 1535
Bit 207
Bit 1145
Bit 771
0.085
0.062
0.058
0.042
0.037
0.034
0.033
0.031
0.030
0.029
0.028
0.026
0.024
0.024
0.023
Morgan Fingerprint Bit Index
Feature Importance (XGBoost)


Chemical Space & Molecular Property Analysis
Chemical Space: MW vs LogP
LogP (Lipophilicity)
Molecular Weight (Da)
Binding Affinity (kcal/mol)
TPSA vs Binding Affinity
Binding Affinity (kcal/mol)
Topological Polar Surface Area (Ų)
Molecular Weight Distribution
Frequency
Molecular Weight (Da)
Binding Affinity Distribution
Frequency
Binding Affinity (kcal/mol)